

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2003, 07:12:33 ; Search time 4178 Seconds
(without alignments)
3963.498 Million cell updates/sec

Title: US-09-805-337A-2
Perfect score: 3183
Sequence: 1 MLLEFSVILISWVSTVGEG.....SSPPFRALCEKFEYPICE 569

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ -p2n.model -DEV=xih
-O/cgpn2.1/USPTO/US09805337/runal.17072003.135605.18919/app.query.fasta.1.711
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09805337.@CGN.1.1.2586.@runal.17072003.135605.18919 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database: GenEmbl.*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3183	100.0	2823	6 AX274975	AX274975 Sequence
2	3183	100.0	2823	9 AF295327	AF295327 Homo sapi
3	2733	85.9	1707	6 AX274977	AX274977 Sequence
4	1536	48.3	3926	9 HSH	Y00716 Human mRNA
5	1537.5	48.3	3934	4 SSC278470	AJ278470 Sus scrofa
6	1524	47.9	2955	10 MUSCFHRD	M29009 Mouse compl
7	1463.5	46.0	4256	10 RNC320522	AJ320522 Rattus no
8	1455.5	45.7	4229	6 A91883	A91883 Sequence 1
9	1426	44.8	2729	10 AFA36847	AFA36847 Rattus no
10	1419.5	44.6	4300	10 MUSH	M12660 Mouse Cfh 1
11	1307	41.1	2715	6 A91885	A91885 Sequence 3
12	1284	40.3	2132	9 HUMCHP	M17517 Human compl
13	1054.5	33.1	1264	9 HUMHAAB	M65293 Human facto
14	1054.5	33.1	1266	9 BC016755	BC016755 Homo sapi
15	1050.5	33.0	1266	6 AX409739	AX409739 Sequence
16	1050.5	33.0	1266	9 HUMHAAA	M65292 Human facto
17	1038.5	32.6	1148	9 HSH361	X56209 Human mRNA
18	1038.5	32.6	1166	9 HSCOMH362	X56210 H. sapiens m
19	927.5	29.1	1040	6 AX411119	AX411119 Sequence
20	927.5	29.1	1040	9 HSDDESK59	X64877 H. sapiens m
21	927.5	29.1	1075	9 BC022283	BC022283 Homo sapi
22	906.5	28.5	1734	10 MUSCFHRB	M29008 Mouse compl
23	861	27.0	2008	4 BTCPMP	X98697 B. taurus mr
24	783	24.6	2358	10 MUSCAPB	BC030166 Mus muscu
25	782	24.6	2377	10 BC030166	M14057 Human facto
26	764.5	24.0	2171	9 HUMFXIITC	I05730 Sequence 21
27	763.5	24.0	2180	6 T05730	AX411132 Sequence
28	710	22.3	1269	6 AX411132	AX411132 Sequence
29	710	22.3	1269	9 HSDOWN	X68679 H. sapiens
30	683	21.5	155892	9 AL353809	AL353809 Human DNA
31	681.5	21.4	1929	6 AX147475	AX147475 Sequence
32	681	21.4	1315	6 AX411193	AX411193 Sequence
33	681	21.4	1315	9 HSCFHRP4	X98337 H. sapiens m
34	663.5	20.8	3071	10 MUSCFHR82	BC026782 Mus muscu
35	631	19.8	2356	10 MUSCFHRC	M29010 Mouse compl
36	561	17.6	1532	6 A91886	A91886 Sequence 4
37	560.5	17.6	1532	6 BC029390	BC029390 Homo sapi
38	522.5	16.4	83166	9 AL139418	AL139418 Human DNA
39	488.5	15.3	755	9 HUMHAAC	M65294 Human facto
40	476	15.0	3377	5 PBXCOMPREG	I21703 Paralabrax
41	460.5	14.5	186760	2 AC026056	AC026056 Homo sapi
42	449	14.1	150626	9 HSI177P10	AL049744 Human DNA
43	449	14.1	180155	2 AL161735	AL161735 Homo sapi
44	437	13.7	73320	9 HSD15D12	AL049741 Human DNA
45	435.5	13.7	4020	4 AF038131	AF038131 Ovis arie

ALIGNMENTS

RESULT 1

AX274975
LOCUS AX274975 2823 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0168695.
ACCESSION AX274975
VERSION AX274975.1 GI:16547598
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 Murphy, B.F.
A novel factor-h related protein 5 and antibodies thereto
Patent: WO 0168695-A 1 20-SEP-2001;
Baxter International Inc. (US)
FEATURES
source 1. 2823
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 960 a 531 c 346 g 786 t
ORIGIN
Alignment Scores:
Pred. No.: 5,49e-276 Length: 2823
Score: 3183.00 Matches: 569
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-805-337A-2 (1-569) x AX274975 (1-2823)
QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
DB 94 AGTGTGCTCTTATTCAGTGTATCCATCATCGGATCCACCTGGGAGCAAGGA 153
QY 21 ThrLeuCysAspPheProIleHisHisGlyPheLeuTrpAsnGluGluAspTrpAsn 40
DB 154 ACACCTTGTGATTTCCAAATAATACACCGATGATTCGTATGATGAACAACATATATAC 213
QY 41 ProPheSerGlnValProThrGlyGluValPheTrpTrpSerCysGluTrpAsnPheVal 60
DB 214 CTTTTTCCAGATTCTCTACAGGGAGTTTCTATTACTCTCTGGAATATATTTTGG 273
QY 61 SerProSerLysSerPheTrpThrArgIleThrCysTrpGluGluGlyTrpSerProThr 80
DB 274 TCTCTTCAAAATCCTTTTGGACGTCGATACATGACAGAGAAGATGCTCACCAACA 333
QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100
DB 334 CCGAAGTGTCTCAGATGCTGTTCTCTTTGTGAAAATGTCATTCGAATCTTCA 393
QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu 120
DB 394 GGAGTAATATCATCTGGAAGGTATCTGATCAAAATTTATTTGCACACAGATATACGCTT 453
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProIleCys 140
DB 454 CAAAACATGACAAAACATTTTCGTGTAGAACGGCGCTGCTCCACATCTCCCATATGC 513
QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluValAsnValAspIleGlnPro 160
DB 514 ACCCTTCACCTAAAGAGAAATGTCATGCTCAATTTTAGAACCAATCTACATCTCAGCCA 573
QY 161 LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
DB 574 AAAAAAGAAAGCTACAAAGTTGAGACCGGTTGAATTTCTCTCGACAGAAAAATCTTATA 633
QY 181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr 200
DB 634 AGAGTTGATCAGACTCATGATCAATGTTACCAATTTGGGTGTCACCTAACCTTCCACA 693
QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyLysValLys 220

DB 694 TCCAAAGCAGACATGATCATGTGCTCCACCTCTCCAACTCCCAATCGTGAACCTTAAG 753
QY 221 GlnIleArgLysGluGlyTrpGlyHisAsnGluValGlnTrpAspCysAsnProAsn 240
DB 754 GAGATAGAGAAAGAGAAATATGACACCAATYGAAGTAGTCGATATGATTCATCTTAAT 813
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGlyTrpThrLeuPro 260
DB 814 TTTATATATTAAGGGCTTAGAATAATACATGCTGTGATGAGATGCAACACTTTACC 873
QY 261 ThrCysValGluGlnValLysThrCysGlyTrpIleProGluLeuGluTrpGlyTrpVal 280
DB 874 ACTTGTGTTGAACAAGTGAACCAATGATGATCAACCTGCAACCTGACGATATGCT 933
QY 281 GlnProSerValProProTrpGlnHisGlyValSerValGlnValAsnCysArgAsnGlu 300
DB 934 CAGCGCTGCTGCTCTCCCTATACATGAGATGTTTACGTGAGGTAAATTCGCAAAATGNA 993
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320
DB 994 TATGCAATGATGGAATATACATGATGATACCTGTATTAATGCAATATGACAGACTCTCT 1053
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
DB 1054 ATGTGTGTTGCACACCACTTAAGAGGTCAAAATAGCAGGATTAAATTAATAAACCA 1113
QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTrpArgCysSer 360
DB 1114 TTACTCAGACTATCGGAGAAACAAATTATCATCTAATTCAGAAATCGTTACGATCTTCA 1173
QY 361 AspIlePheArgTrpArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
DB 1174 GACATCTTCAGATPACAGGCACTCACTCTATTAACCGGAATGCAATCTCAACTAGAC 1233
QY 381 CysTrpGluLysArgGluGlnPheCysProProProGlnIleProAsnIleGlnAsn 400
DB 1234 TGCACAGAAAAGGCAACAAATTCGCCACCGCCACCTCAGATACCTATATGCTCAGANT 1293
QY 401 MetThrThrThrValAsnTrpGlnAspGlyLysValAlaValIleCysLysGluAsn 420
DB 1294 ATGACAAACACAGCTGAATATATCAGATGAGAAAAAGTAGCTCTCTGTAAACAAAMC 1353
QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro 440
DB 1354 TATCTACTTCCAGAAAGCAAAATGTATCTAAAGATGACGATGCCAATCATATTACA 1413
QY 441 ArgCysValGluSerThrAlaTrpCysGlyProProProSerIleAsnAsnGlyAspThr 460
DB 1414 CCCTGTGTGAGTCTACTGCAATATTGTGGGCCCTCTCCATCTATTAAACATGGAATACC 1473
QY 461 ThrSerPheProLeuSerValTrpProProGlySerThrValThrTrpArgCysGlnSer 480
DB 1474 ACCCTATTCCTCATTTATGATATATCTCCAGGGTCACAGTACCTACGTTGCCAGTCC 1533
QY 481 PheTrpLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500
DB 1534 TTTCTATTAACCTCCAGGCTCTCTAATCTAATCATCAACAAATAAACACGTGGTCACAAACA 1593
QY 501 ProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln 520
DB 1594 CCAAGATGCTTAGATCCATGTTGTGTATCTGAAGAAAACATGAAACAAAATATACCTACAG 1653
QY 521 LeuLysTrpArgAsnAspGlyLysLeuTrpAlaLysThrGlyAspAlaValGluPheGln 540
DB 1654 TTTAAATAGAGAAAGAGTGAAGAAACCTATGAGAAAACAGGGAGTCTGTTGAATTCAG 1713
QY 541 CysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560
DB 1714 TGTAAATTTCCCATAAAGCATGATATATCACCACCATTTTCGAGCAATCTGTACAGAA 1773
QY 561 GlyLysPheGlnTrpProIleCysGlu 569
|||||

OY		421	YYRLLEULEUPROGLIAHLVLSYLULVLAIVLYSLVASPGIYAQTPLGINSERLEU	PRO	444
.DB		1354	TATCTACTTCCGAAGACCAAGAATAATTGTATGTAAAGATGCACGATGCCAATCATTACCA	141	
OY		441	ArGcysValGlSerThrAlaTrcCysGlyProProPseTLeasAsnGlyAspThr	460	
.DB		1414	CGCTGTGGTAAGTACTACTGCATATTTGTGGCCCCCTCCATCTATTAAACAATGGACATACC	147	
OY		461	ThrsSerPheProLeuSerValTyrrProPoglyserThrValThrtTyrrArgcysGlnSer	480	
.DB		1474	ACCTATATTCACATTAATCATATATCTCCAGGGTCACACGTAACGTAACCGTTGCCAGTCC	153	
OY		481	PheTYrLysLeuGInGlySerValThryValThrcysArgAsnLysGlnTrpSerGluPro	500	
.DB		1534	TTCTTAAACCTCCAGGGCTCTGTACTGTAACTGTAACTGCAGAAATAAACAGTGTCAGAACCA	159	
OY		501	ProHArgcysLeuAspProcysValValSerGIuGluAsnMetAsnLysAsnAnileGln	520	
.DB		1594	CCMAATGCCCAGAACCCATGTGTGTCTGAAAGAAAACATACACAAAATAACATACACG	165	
OY		521	LeuLTSTPrArgAsnAPolyLysLeuTYrAlaLysThyGlyAspAlaValGluPheGln	540	
.DB		1654	TTAAATATGGAGAAACAGATGGAAACTCTATGCAAAAAACAGGGGATGCTTTGAATTCCAG	171	
OY		541	CysLysPheProHisLysAlaMetLleSerSerProProPheArgAlaLecysGlnGlu	560	
.DB		1714	TGTAAATTCGCCATTAAGCGATGTATCATCATCACCACTTCGAGCAATCTGTCCAGAA	177	
OY		561	GlyLysPheGluTyrrProLecysGlu 569		
.DB		1774	GCGAAATTGGAATATCTATATGTCGAA 1800		
RESULT 3					
AX274977			1707 bp	DNA	linear
LOCUS					
DEFINITION			Sequence 3 from Patent WO0168695.		
ACCESSION			AX274977		
VERSION			AX274977.1		GI:16547599
KEYWORDS					
SOURCE			unidentified.		
ORGANISM			unidentified.		
REFERENCE			unclassified.		
AUTHORS			1		
TITLE			Murphy,B.F.		
JOURNAL			A novel factor-h related protein 5 and antibodies thereto		
FEATURES			Baxter International Inc. (US)		
source			Location/Qualifiers		
			1..1707		
			/organism="unidentified"		
			/db_xref="taxon:32644"		
			1..1707		
misc_feature			/note="Generic sequence"		
BASE COUNT			340 a 170 c 280 g 231 t 686 others		
ORIGIN					
Alignment Scores:					
Pred. No.:			7.49e-236		Length: 1707
Score:			2733.00		Matches: 475
Percent Similarity:			83.48%		Conservative: 0
Best Local Similarity:			83.48%		Mismatches: 94
Query Match:			85.86%		Indels: 0
.DB:			6		Gaps: 0
US-805-805-337A-2 (1-569) x AX274977 (1-1707)					
OY		1	MetLeuLeuPheSerValIleLeuLieserTPValSerThrValGlyGlyGluGly	20	
.DB		1	ATGYTNNTNYNTTWMSGNTNATHYTNDTHMSNTGCGTNMWSNACNGTNGNGNARGSN	60	
OY		21	ThrLeucysAsppheProLysLleHisIsGlyPheLeuTYrAspGluGluAspTryrAsn	40	
.DB		61	ACNTNTTGIGAYTTYCCNAARTRHCAYAGGNTTYTNTAAGYGARGARCAVTAYAY	120	

OY	41	ProPheSerGluInValProThrGlyGluValIlePheTyrTrpSerCysGluIuTyrrAsnIleVal	60
Dd	121	CcNTTYNSCARGTCCNACNGCGNCARGCTNTTATATATASMTSGYCARTAAAYATTYTGIN	180
OY	61	SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGluIuTyrrSerProThr	80
Dd	181	WSNCCMSNAARWSTNTTGTGCACNMGMNATHACNTGTACAGARGARGCNTGMSNCCNACN	240
OY	81	ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer	100
Dd	241	CCNAARCTGYTNMGNAATGTGYSNNTTYCCNTTYTGTAATAAAGCNATYSNCAKSRMSMSN	300
OY	101	GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeu	120
Dd	301	GGNTYNATHACAAYTNGARGCGNCAVACMGTNCNCAATHTATGTGAATAAACNGNTAAYSNTYN	360
OY	121	GlnAsnAsnGluLysAsnIleSerCysValGluIuArgGlyTrpSerThrProProIleCys	140
Dd	361	CARAAYAAAYGAAAATAAATHTSMSTGTCNGARMGNGTGTGMSNACCNCNNCATHTGY	420
OY	141	SerPheThrLysGlyGluCysHisValProIleLeuGlnAlaAsnValAspAlaGlnPro	160
Dd	421	WSNTTTYACNAARGNGARTGYCATGTGCCNATHYTNARGCANATYTNCATGECNCRCCN	480
OY	161	LysLysGluSerTyrLysValGlyLysPylLeuLysPheSerCysArgLysAsnLeuIle	180
Dd	481	AARAARGARWSTNTAAARCTGNCGNCAVGTNYTNAARTTYTSMTGYGNAARAAYTNTATH	540
OY	181	ArgValGlySerAspSerValGlnCysTyrGlnIlePheGlyTrpSerProAsnPheProThr	200
Dd	541	MONGTNGCMSGAIWSTNCTNCAFTGCTYACARTTTGGTGTGMSNCCNAATTTTCCNACN	600
OY	201	CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyLysValLys	220
Dd	601	TGYAARGCNCAARGTWMGMSNTGYGNCNCCNCCNCAATXTMNSNAAYGNGARGTNAAR	660
OY	221	GluIleArgLysGluGluGlyGlnHisAsnGluValValGluTyrAspCysAsnProAsn	240
Dd	661	GARATHMGNAAGAARGARTAYGCGNCATYAAYGARGTNGTNGARTTAAGATYGAYGCCMAAY	720
OY	241	PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGlyTrpThrIleLeuPro	260
Dd	721	TTYATHATHAAAYGCMCCNARAARAATHCARTGTGTGNGATGAGNGARTGCMCMNTYTNCCN	780
OY	261	ThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluIuTyrrGlyTyrVal	280
Dd	781	ACNTGTGTNGARCACGTNAARACNTGTGCGTNTATTHCCNGARTYTNCAARTAYGCGNTATGN	840
OY	281	GlnProSerValProProProGlnHisGlyValSerValGluValAsnCysArgAsnGlu	300
Dd	841	CARCCNMSGNCCNCCNATYACARAYGGNGTNGMSNGTNGARTNAAYTGYMGNAAYGAR	900
OY	301	TyralaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro	320
Dd	901	TAYGCMATGATHGCGNAAYAAATATGATHAMCMTGTATHAAYAGMATHTHGCACNCAARTYTNCCN	960
OY	321	MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr	340
Dd	961	ATGTGCTGTCNCCNACCAVCARTTNAARBMGTNGYAAARATGCGCGNGCTNAAYATHTHAARCN	1020
OY	341	LeuLeuLysLeuSerGlyLysGluLysPheAsnHisAsnSerArgIleArgTyrArgCysSer	360
Dd	1021	YTNTNTNAARITYTNMSGNABARGARTTYAACYYAAATYSMSGMATHTMGNTATYAGNTCYWSN	1080
OY	361	AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp	380
Dd	1081	GAYATHHTTYGMTNTFAWMCAMCAWSTNCTNGYATHAAAYGCAAAATGCAAYCCNCAARGTNGAY	1140
OY	381	CysThrGlyLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn	400
Dd	1141	TGYACNGRAARBMGARGARTTYTGYCCNCCNCCNCCCAATHTCCNAAYCCNCCARAY	1200

Qy 401 MetThrThrValAsnTyrGlnAspGlyGluValAlaValLeuCysIleGluAsn 420
|||||
Db 1201 ATGACNACNAGCTTAATATACAGATGNGARAAAGTGTNTGTGTAARAGARAY 1260
Qy 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTTrpGlnSerLeuPro 440
|||||
Db 1261 TAYTNTNCCGAGCAGCAARAGATGTCNTGTAARAGAYGGMGTGCGARMSNTYTCN 1320
Qy 441 ArgCysValGluSerThrAlaTyrCysGlyProProPheSerIleAsnAsnGlyAspThr 460
|||||
Db 1321 MGNITGTCNGARMSNACNACNTATGTGCGNCCNCCNCCMSNATTAATATATGNGATYACN 1380
Qy 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
|||||
Db 1381 ACNMSNTTYCCNVTNMSNCTATACCCNCCNGNMSNACNTACNMTATYMGNTGYCARMSN 1440
Qy 481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTTrpSerGluPro 500
|||||
Db 1441 TTYTAYAAATYTCAGAGMNSNCTNACNCTGTGMAAAYAAACARTGMSNARCCN 1500
Qy 501 ProArgCysLeuAspProCysValAlaSerGluGluAsnMetAsnLysAsnIleGln 520
|||||
Db 1501 CCNMGNTGTYTNGATACCTNTGTGTNTGNSNGARAGAAATATGATAAATATATATTCAR 1560
Qy 521 LeuLysTTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540
|||||
Db 1561 YTNAAATGTGMAAAYGAGVAGNAAATYNTATGACNAAACNGAGVAGCNGTNGARTTYCAR 1620
Qy 541 CysLysPheProHISLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560
|||||
Db 1621 TGTAAATTYCCCAATAAACGCAATGATHMSNMSNCCNCTTYMGNCNATHTGTCARGAR 1680
Qy 561 GlyLysPheGluTyrProIleCysGlu 569
|||||
Db 1681 CGNAARTTYGARTATCCNATHTGTCAR 1707

RESULT 4
HSH
LOCUS HSH 3926 bp mRNA linear PRI 08-APR-1997
DEFINITION Human mRNA for complement factor H.
ACCESSION Y00716 M32093 X07525
VERSION Y00716.1 GI:31964
KEYWORDS complement factor H; glycoprotein; regulatory protein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ripoché,J., Day,A.J., Harris,T.J. and Slim,R.B.
TITLE The complete amino acid sequence of human complement factor H
JOURNAL Biochem. J. 249 (2), 593-602 (1988)
MEDLINE 88134059
REFERENCE 2 (bases 1 to 3926)
AUTHORS Day,A.J.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-1988) Day A.J., MRC Immunochemistry Unit, Dept.
of Biochemistry, University of Oxford, South Parks Road, Oxford,
OX1 3OU
COMMENT On Jul 25, 2002 this sequence version replaced gi:183745.
see Immunogenetics 27:211-214(1988) concerning Tyr <-> His at AA
384
see X07523 for truncated (1.8 kb) mRNA sequence.
FEATURES
source
1..3926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="B-38-1, R2a and PE3"
/tissue_type="liver"
/clone_lib="human liver cDNA in pGL153/PvuII/8"
74..3769
/codon_start=1
/product="factor H"
/protein_id="CAA68704.1"

/db_xref="GI:31965"
/db_xref="SWISS-PROT:P08603"
/translation="MRLAKITCLIMLMAICVADENELPPRENTIELTSGMSDQTYPE
GTQATYKCRPGTGRISLGNVIMVCKKEMVALNLRKQCRPCGHDPDTPGTTLTGGN
VFEVGNKAVYTCNEGTYQLNGETINRECDTGNIDNIPICEVVKCLPVAPENKGLVSS
AMEPDRHEFGQAVFVGVCSYKRIEGDEMHSDGFWMSKEVKCEIPLVSCSKSPDIVNG
SPISQRIYKENERFOYCKMNGEYSERDADTESGMPLPSCEKESCDNYIINGD
YSPLRIRKTRGDEITQCCNGMYPATRGATACCTGTGMPARCTLKPCDDYDIDHGG
LYENHMRPRFPVAVGKYSYVCDHPFESGSDYHCHODDGSPPATRCIKRKYFO
VLENENONHGRKAFVQGSIDYACRPGVALPRAQTTVTCMENMSPTPRCIVKTKCSK
STIDLENGFTISQYTLAKKAKYQCKLGYTAGETSGSIRCGKDGSAOPTIKS
CDIPEVMARNTDNFTKRLNDTLDYEDHGESVTGTSIYCGVMWSLPLCYE
RECELPKIDVHLVPRKTDQYKGVLEFSCPKFTIYGPNSVQCYHFGIDSLPICK
BOVQSGPPELLNGNVEKTEKEEGSHSVAYCYCNPRFLMKGPRIQCVDEGMTLP
VCTIEPSTNGDIPLEHGMALSSPPRYAGDSVEFNCSSFTMIGRSTCTHGYWTO
LPQCAIDKLKCKKSSNLIILEHLLKPKKEPNSNIRRCRKGKGIHTYVINKRMD
PEVNCSSMOIOLCPPPOIPNSHNNNTYTNVNDGKRVSVLCENYLLIDGEEITCKDG
RMQSIPLCEKELPCSQPOLIEGHTINSSSHQVHNTKRLYKTCGEGFRISENETT
CYMGKMSHPQCEGLPKSPPEISGVVAHMSDSYQGEVYKCFEGFGIDSPAIK
CLGKMSHPSCIKTIDLSLPSFENAPIPEKEDYKKAQEVYTCATYKMKDGSNV
TCINSRMTGRPTCRPTSCYNPTVONAYIVSRQMSKYPGSEVRYOCSPYMPQDE
VMCLNGNTPERQCDSTGKCCPPPIDNGDITSPPLSYATARASSVEYOCOLYDLEG
MKRTICRMQMSPPKCLPVCYISREIMENTIALRMWAKOKLYSRFGSEVFEVCKRG
YRLSSRSHTLRTTCMDKLEYPCKAKR"
74..127
128..3766
/product="factor H"
994
variation
/note="polymorphism"
/replace="c"
1277
/note="polymorphism"
/replace="t"
1409
misc_feature
/note="divergence point of full length (4.3 kb) and
truncated (1.8 kb) mRNA species"
2881
variation
/note="polymorphism"
/replace="g"
BASE COUNT 1306 a 689 c 855 g 1076 t
ORIGIN
Alignment Scores:
Pred. No.: 5,18e-128 Length: 3926
Score: 1538.00 Matches: 298
Percent Similarity: 43.25% Conservative: 93
Best Local Similarity: 32.96% Mismatches: 155
Query Match: 48.32% Indels: 358
Gaps: 3
US-09-805-337A-2 (1-569) x HSH (1-3926)
Qy 23 CysAspPheProLysIleHisIleGlyPheLeuTyrAspGluGluAspTyrAsnProPhe 42
|||||
Db 1046 TGTGATATACAGACATTAACATGAGGCTATATACAGAGAAATATGCGTAGACCATAC 1105
Qy 43 SerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheValSerPro 62
|||||
Db 1106 TTTCAGTAGACGTAGAGAAATATTACTCTCTATTACTGTGATGAGACATTTTGAGACTCGG 1165
Qy 63 SerLysSerPheThrPThrArgGlyLeuThrCysThrGlnGluGlyTTrpSerProThrProLys 82
|||||
Db 1166 TCAGGAGTACTAGGATGACATTCATTCACACAAAGATGATGTCGCGACAGATACCA 1225
Qy 83 CysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGlnSerGlyLeu 102
|||||
Db 1226 TCCCTCAGAAATGTATTTCTTATTTGGAATAATGATATATACAAATATATGGAACA 1285
Qy 103 IleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGlnAsn 122
|||||
Db 1286 AAGTTTGTACAGGTAATCTATAGACGTTGCGCTGCAATCGGCTTCCCAAAA 1345
Qy 123 AsnGluLysAsnIleSerCysValGluArgGlyTTrpSerThrProProIleCys----- 140

```

Db      1346 GCGCAGACACCACTTACATGTATGAGAAWGGCTGCTCTACTCCAGATGCAATCCGT 1405
OY      140 -----
Db      1406 GTCAAAACATGTTCCAAATCAAGTATAGATATGAGAAATGGGTTATTTCTGAATCTCAG 1465
OY      140 -----
Db      1466 TATACATATGCTTAAAGAAAAAGGAATATCAATGCAACTAGCATATGTAACAGCA 1525
OY      140 -----
Db      1526 GATGTTGAACATCAGAGATCAATTAGATGTGGAAAGATGATGTCAGCTCAGCCACAG 1585
OY      141 -----
Db      1586 TGCATTAAATCTTGTGATATCCAGATTTATGAATGCCAGAACTAAAAATGACTTCACA 1645
OY      143 -----
Db      1646 TGGTTTAAGCTGAATGACACATTTGACATGATGCCATGATGTTATGAAAGCAATACT 1705
OY      143 -----
Db      1706 GGAAGCACCACGTGTTCCATAGTGTGTGTTACAAATGTTGGTCTGATTTACCCATATGT 1765
OY      144 ---LysGluGluCysHisValProIleLeuGluAlaAsnValAspIleGlnProLysLys 162
Db      1766 TATGAAGAAGATGCGAACTTCTCTAAATAGATGTAACACTTATGTCCTCATGCGCAAGAA 1825
OY      163 GluSerTyrLysValGluYasPvalLeuLysPheSerCysArgLysAsnLeuIleArgVal 182
Db      1826 GACCAATAAAGATTGAGAGGCTTGGAATTTCTCTCCCAACCGAGATTTCAAATATGTT 1885
OY      183 GlySerAspSerValGlnCysTyrGlnPheGlyTyrPserProAsnPheProThrCysLys 202
Db      1886 GGAACCTAAATTCGTTGAGTGTACCACTTGGATTTGTCCTGACCTCCCAATATGTAATA 1945
OY      203 GlnGlnValArgSerCysGlyProProGlnLeuSerAsnGlyValLysGluLeu 222
Db      1946 GAGCAAGTACATCATGTCGTGCACCTCCTGAACCTCCTCAATGGGAATTTAAGAAATA 2005
OY      223 ArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsnPheIle 242
Db      2006 ACGAAAGAAATATGACACAGTGAAGTGTGAATATATTATTCATCTGATTTCTA 2065
OY      243 IleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrLeuProThrCys 262
Db      2066 ATGAAAGGACCTAATAAATTTCAATGTGTGATGGAGGTGACACACTTTACCAAGTGT 2125
OY      263 ValGluGlnValLysThrCysGlyTyrIleProGlnLeuGluTyrGlyValGlnPro 282
Db      2126 ATTTGGAGAGAGATGACCTGTGAGATATACCTGAACCTTGACATGCTGGGCCGAGCT 2185
OY      283 SerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnLysIleAla 302
Db      2186 TCTTCCCTCTTACTATGAGATTCAGTGAATTCGAATTCGAATTCAGCAATCATTTTACA 2245
OY      303 MetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThrGluLeuProMetCys 322
Db      2246 ATGATTTGACACAGATCAATTTACGTATTTCAATGAGATGACCACTTCCCACTGT 2305
OY      323 ValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeuLeu 342
Db      2306 GTGGCAATGATAACTTAAGAAGTGCAAATCATCAAAATTAATATACTTACAGAACAT 2365
OY      343 LysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAspIle 362
Db      2366 TTAATAAACAAAGAGAAATTCATATATTCATCAATTAAGGTACAGATGTGAGAGAAAA 2425
OY      363 PheArgTyrArgHisSerValCysIleAsnGlyLysTyrPasnProGlnValAspCysThr 382

```

```

Db      2426 GAAGATGATACACACAGCTGTGATTAATGAGAGATGGATCCAGAACTGACTCTCA 2485
OY      383 GluLysArgGluGlnPheCysProProGlnIleProAsnAlaGlnAsnMetThr 402
Db      2486 ATGGCACAATATCAATTTATGCTCCACCTCCACCTCAGATTTCCCAATTTCCACATATGACA 2545
OY      403 ThrThrValAsnTyrGlnHisPglLysLysValAlaValLeuCysLysHisAsnTyrLeu 422
Db      2546 ACCACACTGAATTTATCGGATGAGAGAAAAAGTATCTGTTCTTGGCAAGAAATATATCTA 2605
OY      423 LeuProGluAlaLysGluIleValCysLysAspGlyArgTyrPheGlnSerLeuProArgCys 442
Db      2606 ATTCAGAGAGAGAGAAATTTACATGCAAAAGTGAAGATGGAGTCAATACCACTCTGT 2665
OY      443 ValGlu 444
Db      2666 GTTGAATAAATTCATGTTCCACACACCTCAGATTAAGAACGGAACCTTAATTCATCC 2725
OY      444 -----
Db      2726 AGGTCTCACAAGAAAGTTATGACATGGAGCTAAATGATTAATCTTGTGAGGCTGT 2785
OY      444 -----
Db      2786 TTCCAGATATCTGAAGAAATTAAGAACACATGCTCATGCGAAATGCACTTCTCCACCT 2845
OY      444 -----
Db      2846 CAGTGTGAAGGCTTCTCTGTAATCTCCACCTGAGATTTCTCATGCTGTGTACTCTCAC 2905
OY      444 -----
Db      2906 ATGACAGACGTTATCAGTATGAGAGAGAAGTTAGTACAAATGTTTGAAGTTTGA 2965
OY      444 -----
Db      2966 ATTTGATGGCTTCGCAATTTGCAAAATGCTTAGAGAAAAATGCTTCACCTTCATCATGC 3025
OY      444 -----
Db      3026 ATAAAGAGATTGTCTCTAGTTTACCTACCTTTGAATAATGCCATACCATGGAGAGAA 3085
OY      444 -----
Db      3086 AAGATGTGTATTAAGCGGGTGAGAGAGTACCTTACACTTGTGCAACATATTACAATAATG 3145
OY      444 -----
Db      3146 GATGAGCCAGTAAATGTAATGATGATTAATAGCAGATGACAGAGAAAGCCAAATGCACA 3205
OY      444 -----
Db      3206 GACACCTCTGTGTGAATCCGCCACAGTACAATAATGTTATATAGTGCAGACAGATG 3265
OY      444 -----
Db      3266 ACTAAATATCATCTGTGTGAGAGAGTACGTTATCAATGTAGAGCCCTTATGCAAAATGTTT 3325
OY      444 -----
Db      3326 GGGGATGAAGAGTATGTTTAATGGAATCGAGGAGAACCTGAGCAACCACTCAATGCAAAAGT 3385
OY      445 SerThrAlaTyrCysGlyProProSerIleAsnAsnGlyAspThrThrSerPhePro 464
Db      3386 TCTACAGCAAAATGTTGGGCCCTCCACTTATTTGACAAATGGGAGATTAATTCATTTCCCG 3445
OY      465 LeuSerValTyrProProGlySerThrValIleTyrArgCysGlnSerPheTyrLysLeu 484
Db      3446 TTGTAGATATATGCTCCACCTTCATCAGTGTAGTACCAATGCCAAACTGTGATCAACT 3505
OY      485 GlnGlySerValThrValThrCysArgAsnLysGlnTyrPserGlnProArgCysLeu 504
Db      3506 GAGGTTAACAGCAGATTAATGATGTAAGAAATGACATGTGTGCAAGCAACCAAAATGCTTA 3565

```

OY		505	AsprocyvalValasIserglugluuasmecnaslnysnknsllegInleuylstrparg	524
			
Db		3666	CATCCGGTGTGAATAATCCCAGAAATTATTAACCAATGAAAGTTGGTGACA	3625
OY		525	AsnapsglylsLeuTyValatylThrgJyAspaIalValgIuphegnCysLysPhePro	544
			
Db		3686	GCCAAACAGAAGCTTTATTCGACAGACGATCATGTTGATTTGTCTGTAACCGGGA	3685
OY		545	HisllysAlametiLeserSerproPhearghAalIelcysglngluglYlvsphegl	564
			:::::	
Db		3686	TATGCTCTTCATCACGCTTCACACATTCGCAGAACACATCTTGATGGAAACTGGAG	3745
OY		565	TyrProIleCys	568
Db		3746	TATCCAACCTGT	3757
RESULT 5				
LOCUS		SSC278470		
DEFINITION		SSC278470	3934 bp mRNA linear MAM 27-JUN-2002	
ACCESSION		AJ278470	Sus scrofa mRNA for putative porcine complement factor H (PFH-11 gene)	
VERSION		AJ278470.1	GI:21464593	
KEYWORDS		PFH-11 gene; porcine complement factor H.		
SOURCE		Sus scrofa		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.		
REFERENCE		1	Hegasy,G., Willhoeft,U., Manjo,S., Zipfel,P.F. and Hellwage,J.	
AUTHORS			Porcine complement regulator factor H: functional cloning,	
TITLE			functional characterization and relevance to cross-species transplantation	
JOURNAL			Unpublished	
REFERENCE		2	(bases 1 to 3934)	
AUTHORS			Hegasy,G.	
TITLE			Direct Submission	
JOURNAL			Submitted (16-JUN-2000) Hegasy G., Research Group of Biomolecular Medicine, Bernhard Nocht Institute, Bernhard-Nocht-Str. 74, Hamburg D-20359, GERMANY	
FEATURES			Location/Qualifiers	
source			1..3934	
			/organism="Sus scrofa"	
			/db_xref="taxon:9823"	
			/clone="PFH-11"	
			/sex="female"	
			/tissue_type="liver"	
			/dev_stage="adult"	
			/tissue_lib="porcine liver unizap ghl"	
			/country="Germany"	
			1..3705	
			/gene="PFH-11"	
			1..3705	
			/gene="pfh-11"	
			/function="complement regulatory protein"	
			/codon_start=1	
			/product="putative porcine complement factor H"	
			/protein_id="CAC81999.1"	
			/db_xref="GI:21464594"	
			/translation="MRCPAKFWMLILMTVCVAEDCRREPPRKETELVSGSWTEQTPPE GTQATYRCRPGRFTIGSLIVMYCKDKGNWLSHPSTICRRKPCGHGPDTSFGSHLATVGPTE OFEAGAVVAYDCDEGYOMLEINRECDNTMNINPLCEVKLKLPTEPNHGIVSG ALFDDQATVOOVYOCOMNGSYSDGRKHSCGFCGCVSMGPKPCVGISCTPRNIMG YPTISKEITRYKNENVOYKCSAGEFYSERGETICTASGMARVPSCVELTCDPIHPNFQ YTPSNRIRTRDRIITYHCESGEYPEIDGNNAKRGIMNSVAPRCTLKPCSAPYIKRHKL LYDYRGYPFAKNVGQYFYVYCDHNHFVSRLSGDYLCTCKRMGSMAEYPCLOCFEDVL EHGNYSYSAKKYLGEGTVRVOCYGYSLIPMODILLCTENMGSPKCIIVKTCAKSD IEINGFPSENIIFYPLNKOTORYCKRGYVADGATSGLIITCGMSARPVCIKSDD AILENNARIKSDGTWFLINDYVDECOOGFSRSGDAGSTVCGEDMSDKPACYERK CTPGIKSYLDAEKRHKFKYGVYLAKHCHERRMLGADSVOCTHFCSMPVLPYCAG QVSKCAPPOLLINEGVKETQGEYOHEVEVYCNPRLMKGSHKIQCVDEWTTLPEV CIEBERKCQDIPDLDHGVYRSAPYEGHDSVERGCAGNASFSLVGRSTICISGMTOLD	
gene				
CDS				

sig_peptide	/gene="pfh 11" 55..3702 /product="porcine complement factor H" join(1..34,55..3702) /gene="pfh 11" 55..240 /rpt_family="short consensus repeat (SCR) #1 [or CCP or sushi unit]" /rpt_unit=56..240 241..426 /rpt_family="SCR #2" /rpt_unit=241..426 427..618 /rpt_family="SCR #3" /rpt_unit=427..618 619..789 /rpt_family="SCR #4" /rpt_unit=619..789 790..963 /rpt_family="SCR #5" 964..1152 /rpt_family="SCR #6" 1153..1329 /rpt_family="SCR #7" 1330..1512 /rpt_family="SCR #8" 1513..1689 /rpt_family="SCR #9" 1690..1866 /rpt_family="SCR #10" 1867..2052 /rpt_family="SCR #11" 2053..2232 /rpt_family="SCR #12" 2233..2409 /rpt_family="SCR #13" 2410..2592 /rpt_family="SCR #14" 2593..2799 /rpt_family="SCR #15" 2800..2973 /rpt_family="SCR #16" 2974..3150 /rpt_family="SCR #17" 3151..3324 /rpt_family="SCR #18" 3325..3507 /rpt_family="SCR #19" 3508..3702 /rpt_family="SCR #20" 3706..3924 3924..3934			POCAATGKIKKCAKSLTFAAGNLTLDKTEPDHNTNOSTKRGSKSEKSYSCINJCMWPR RVSKKEEVLNSCPPPPOTPNAAQDMTTVTNKDKGKISILCOENIT1DDGEIYKDKGR WQSPVRCVERKIPCESEPEIDGTTIOSPGSAEREMEPKKYAGTFLSYCEGEPFEL SEKTI1CHMGKMSLPOCVGRPCAPPEIONGAAOBRDXYLGEVATYSCDEBFGID GPAICRLIGEKMSHPPECINTDCFLPTFNDAVAVIGPKPSYKSGEVTFCILOFOD EGVATRICIKSMVIGKPCACRVCSCVNPSPVNAVILDKRSRYOGERARYCICPYAL YGDVEMCFNGSMTRPPROCKDKGKCGPPPIIDNGDTSPFLPYPPGTIVYECOCYS YELGSSSIRKCNQMSDEPRCLDACVYSEMMKRNHIELKRPDKILYSTDDTIEF KCRGGYRTPTDLHFRATCOOGKVAVPTCG- 1..54			BASE COUNT 1254 a 896 g 1010 t		
mat_peptide	/gene="pfh 11" 55..3702 /product="porcine complement factor H" join(1..34,55..3702) /gene="pfh 11" 55..240 /rpt_family="short consensus repeat (SCR) #1 [or CCP or sushi unit]" /rpt_unit=56..240 241..426 /rpt_family="SCR #2" /rpt_unit=241..426 427..618 /rpt_family="SCR #3" /rpt_unit=427..618 619..789 /rpt_family="SCR #4" /rpt_unit=619..789 790..963 /rpt_family="SCR #5" 964..1152 /rpt_family="SCR #6" 1153..1329 /rpt_family="SCR #7" 1330..1512 /rpt_family="SCR #8" 1513..1689 /rpt_family="SCR #9" 1690..1866 /rpt_family="SCR #10" 1867..2052 /rpt_family="SCR #11" 2053..2232 /rpt_family="SCR #12" 2233..2409 /rpt_family="SCR #13" 2410..2592 /rpt_family="SCR #14" 2593..2799 /rpt_family="SCR #15" 2800..2973 /rpt_family="SCR #16" 2974..3150 /rpt_family="SCR #17" 3151..3324 /rpt_family="SCR #18" 3325..3507 /rpt_family="SCR #19" 3508..3702 /rpt_family="SCR #20" 3706..3924 3924..3934			POCAATGKIKKCAKSLTFAAGNLTLDKTEPDHNTNOSTKRGSKSEKSYSCINJCMWPR RVSKKEEVLNSCPPPPOTPNAAQDMTTVTNKDKGKISILCOENIT1DDGEIYKDKGR WQSPVRCVERKIPCESEPEIDGTTIOSPGSAEREMEPKKYAGTFLSYCEGEPFEL SEKTI1CHMGKMSLPOCVGRPCAPPEIONGAAOBRDXYLGEVATYSCDEBFGID GPAICRLIGEKMSHPPECINTDCFLPTFNDAVAVIGPKPSYKSGEVTFCILOFOD EGVATRICIKSMVIGKPCACRVCSCVNPSPVNAVILDKRSRYOGERARYCICPYAL YGDVEMCFNGSMTRPPROCKDKGKCGPPPIIDNGDTSPFLPYPPGTIVYECOCYS YELGSSSIRKCNQMSDEPRCLDACVYSEMMKRNHIELKRPDKILYSTDDTIEF KCRGGYRTPTDLHFRATCOOGKVAVPTCG- 1..54			BASE COUNT 1254 a 896 g 1010 t		
gene	/gene="pfh 11" 55..3702 /product="porcine complement factor H" join(1..34,55..3702) /gene="pfh 11" 55..240 /rpt_family="short consensus repeat (SCR) #1 [or CCP or sushi unit]" /rpt_unit=56..240 241..426 /rpt_family="SCR #2" /rpt_unit=241..426 427..618 /rpt_family="SCR #3" /rpt_unit=427..618 619..789 /rpt_family="SCR #4" /rpt_unit=619..789 790..963 /rpt_family="SCR #5" 964..1152 /rpt_family="SCR #6" 1153..1329 /rpt_family="SCR #7" 1330..1512 /rpt_family="SCR #8" 1513..1689 /rpt_family="SCR #9" 1690..1866 /rpt_family="SCR #10" 1867..2052 /rpt_family="SCR #11" 2053..2232 /rpt_family="SCR #12" 2233..2409 /rpt_family="SCR #13" 2410..2592 /rpt_family="SCR #14" 2593..2799 /rpt_family="SCR #15" 2800..2973 /rpt_family="SCR #16" 2974..3150 /rpt_family="SCR #17" 3151..3324 /rpt_family="SCR #18" 3325..3507 /rpt_family="SCR #19" 3508..3702 /rpt_family="SCR #20" 3706..3924 3924..3934			POCAATGKIKKCAKSLTFAAGNLTLDKTEPDHNTNOSTKRGSKSEKSYSCINJCMWPR RVSKKEEVLNSCPPPPOTPNAAQDMTTVTNKDKGKISILCOENIT1DDGEIYKDKGR WQSPVRCVERKIPCESEPEIDGTTIOSPGSAEREMEPKKYAGTFLSYCEGEPFEL SEKTI1CHMGKMSLPOCVGRPCAPPEIONGAAOBRDXYLGEVATYSCDEBFGID GPAICRLIGEKMSHPPECINTDCFLPTFNDAVAVIGPKPSYKSGEVTFCILOFOD EGVATRICIKSMVIGKPCACRVCSCVNPSPVNAVILDKRSRYOGERARYCICPYAL YGDVEMCFNGSMTRPPROCKDKGKCGPPPIIDNGDTSPFLPYPPGTIVYECOCYS YELGSSSIRKCNQMSDEPRCLDACVYSEMMKRNHIELKRPDKILYSTDDTIEF KCRGGYRTPTDLHFRATCOOGKVAVPTCG- 1..54			BASE COUNT 1254 a 896 g 1010 t		
repeat_region	/gene="pfh 11" 55..3702 /product="porcine complement factor H" join(1..34,55..3702) /gene="pfh 11" 55..240 /rpt_family="short consensus repeat (SCR) #1 [or CCP or sushi unit]" /rpt_unit=56..240 241..426 /rpt_family="SCR #2" /rpt_unit=241..426 427..618 /rpt_family="SCR #3" /rpt_unit=427..618 619..789 /rpt_family="SCR #4" /rpt_unit=619..789 790..963 /rpt_family="SCR #5" 964..1152 /rpt_family="SCR #6" 1153..1329 /rpt_family="SCR #7" 1330..1512 /rpt_family="SCR #8" 1513..1689 /rpt_family="SCR #9" 1690..1866 /rpt_family="SCR #10" 1867..2052 /rpt_family="SCR #11" 2053..2232 /rpt_family="SCR #12" 2233..2409 /rpt_family="SCR #13" 2410..2592 /rpt_family="SCR #14" 2593..2799 /rpt_family="SCR #15" 2800..2973 /rpt_family="SCR #16" 2974..3150 /rpt_family="SCR #17" 3151..3324 /rpt_family="SCR #18" 3325..3507 /rpt_family="SCR #19" 3508..3702 /rpt_family="SCR #20" 3706..3924 3924..3934			POCAATGKIKKCAKSLTFAAGNLTLDKTEPDHNTNOSTKRGSKSEKSYSCINJCMWPR RVSKKEEVLNSCPPPPOTPNAAQDMTTVTNKDKGKISILCOENIT1DDGEIYKDKGR WQSPVRCVERKIPCESEPEIDGTTIOSPGSAEREMEPKKYAGTFLSYCEGEPFEL SEKTI1CHMGKMSLPOCVGRPCAPPEIONGAAOBRDXYLGEVATYSCDEBFGID GPAICRLIGEKMSHPPECINTDCFLPTFNDAVAVIGPKPSYKSGEVTFCILOFOD EGVATRICIKSMVIGKPCACRVCSCVNPSPVNAVILDKRSRYOGERARYCICPYAL YGDVEMCFNGSMTRPPROCKDKGKCGPPPIIDNGDTSPFLPYPPGTIVYECOCYS YELGSSSIRKCNQMSDEPRCLDACVYSEMMKRNHIELKRPDKILYSTDDTIEF KCRGGYRTPTDLHFRATCOOGKVAVPTCG- 1..54			BASE COUNT 1254 a 896 g 1010 t		
repeat_region	/gene="pfh 11" 55..3702 /product="porcine complement factor H" join(1..34,55..3702) /gene="pfh 11" 55..240 /rpt_family="short consensus repeat (SCR) #1 [or CCP or sushi unit]" /rpt_unit=56..240 241..426 /rpt_family="SCR #2" /rpt_unit=241..426 427..618 /rpt_family="SCR #3" /rpt_unit=427..618 619..789 /rpt_family="SCR #4" /rpt_unit=619..789 790..963 /rpt_family="SCR #5" 964..1152 /rpt_family="SCR #6" 1153..1329 /rpt_family="SCR #7" 1330..1512 /rpt_family="SCR #8" 1513..1689 /rpt_family="SCR #9" 1690..1866 /rpt_family="SCR #10" 1867..2052 /rpt_family="SCR #1								

US-09-805-337a-2 (1-569) x, SSC278470 (1-3934)

QY 23 CysAspPheProLysIleHisHisGlyPheLeuTyrAspGluGluAspTyrAsnProPhe 42
 Db 973 TGTAGCCGCTCCAGTCTAAACATGGAAGCTATAT-----TATGATCTGTGGATAC 1026
 QY 43 SerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheValSerPro 62
 Db 1027 TTTCAGCAAAATGTAGACAGATATTTTACTACTACTGATCATATTTTGGACTGCC 1086
 QY 63 SerLysSerPheThrPheThrArgIleThrCysThrGluGlyTyrPheProThrProLys 82
 Db 1087 TCACGCCGCTCCGGGACTACCTGACGTCAAAACGGAATGGATGGCTCTCAGAAAGTCCA 1146
 QY 83 CysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerGlyLeu 102
 Db 1147 TGCCTCAGACANTGTATTTTGTATTTTGGAAACATGACACACACATTCATTCAGCAAAA 1206
 QY 103 IleHisLeuGluGlyAspThrValGlnIleLeuCysAsnThrGlyTyrSerLeuGlnAsn 122
 Db 1207 AATATTTTACAGGTGAACCTGTAAAGATTCAAGTTCATCCGCTACAGTCTCCAGAT 1266
 QY 123 AsnGluLysAsnIleSerCysValGluArgGlyTyrPheThrProProIleCysSerPhe 142
 Db 1267 AATCAGACACGCTGCTGTGTACAGAGAAATGGCTGCTCTCTCCCAAGTCATCTTT 1326
 QY 143 ThrLys----- 144
 Db 1327 GTCAAAACATGTGCAGAAATCAGATATAGAAATTGGAAATGGATTTTTTCTCAAAATATA 1386
 QY 144 ----- 144
 Db 1387 TTTACATATCCCTTAATTAACAAACACATATTAAGTGAACAGAGATGTAAACAGCA 1446
 QY 144 ----- 144
 Db 1447 GATGGCGCACTTCAGAGATTAAATCATCTCTGCAAAAGTGATGGTGCAGCTGCAGCTGTG 1506
 QY 144 ----- 144
 Db 1507 TGCATTAACTCTGGATCGCGCACTACCTGGAATGCCAAGATCAAGAGTGTATGTCACAG 1566
 QY 144 ----- 144
 Db 1567 TGTGTTAACTCAATGACACAGTGCATGTGAATGCCACAGAGGTTGAAAGCCGAAT 1626
 QY 144 ----- 144
 Db 1627 GGAGACACTGCAGATCCATTGTGTGTGTGTGAAGACGATGCTGTATAGCCTGCTTGT 1686
 QY 145 -----GlyGluCysHisValProIleLeuGlnAlaAsnValAspAlaGlnProLysLys 162
 Db 1687 TACGAAGAAGATGTATTAATCTCTCGAATAAAAGATATTAGATGCTCAACCAACAAACT 1746
 QY 163 GluSerTyrLysValGlyLysValLeuLysPheSerCysArgLysAsnLeuIleArgVal 182
 Db 1747 AACAAATTCAAAGTTGGAGCGTGTGAATTCACCTGCAGAGAAAGACGTGTGATGCTT 1806
 QY 183 GlySerAspSerValGlnCysTyrGlnPheGlyTyrPheProAsnPheProThrCysLys 202
 Db 1807 GGAGCGATTCAGTTCAATGCTTACCACTTTGGATGGTCCCTTAATCTTCCAAACATGTAA 1866
 QY 203 ---GlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLysGlu 221
 Db 1867 GGAGGGCAAGTAATATCTGTCTCCACCCCTCAACTCTCAATGGGGAATTAAGA 1926
 QY 222 IleArgLysGluGluTyrGlyHisAsnGluValValGluTyrAspCysAsnProAsnPhe 241
 Db 1927 ACACGAAAGAGATATCAACACAGTGGTGTGATACGTTTCAATCTCTGATTTT 1986
 QY 242 IleIleAsnGlyProLysLysIleGlnCysValAspGlyLutTrpThrLeuProThr 261
 Db ----- 261

Db 1987 CTGATGAAGGCTCTCATAAATTCATATGTGTGATGGAATGGAACACTTGGCTGTG 2046
 QY 262 CysValGluGlnValLysThrCysGlyTyrIleProGlnLeuGlnTyrGlyValGln 281
 Db 2047 TGTATTGAGAGAGACGCTGACCTGTGGATATTTCCGACCTGACCATGTGCTGCGG 2106
 QY 282 ProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnLutYr 301
 Db 2107 CATTTGCCCTCCCTCATCATCATCATGAGACTCAGTGGAGTGGGGTGGCAATCATCATTT 2166
 QY 302 AlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThrGluLeuProMet 321
 Db 2167 TCACGTGGTGGACCCAGGTCAATCACCTGTATTAGTGAAGTGAAGTGAAGTTCCTCG 2226
 QY 322 CysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeu 341
 Db 2227 TGCCTTGCACAGATTAACCTCAAGAAAGTAAAGCCTCAAAATATTTTCATCTGAGGA 2286
 QY 342 LeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAsp 361
 Db 2287 AACCTGTGTGATTAAGACCCGAATTTTGACATTAACACCAACCAAAAGTTACAAATGTCGGGA 2346
 QY 362 IlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGluValAspCys 381
 Db 2347 AATTCAGAAAGCAAAATCTCACTCTGTATTAATGGAATGGATGCCAGACTTAAGCTGC 2406
 QY 382 ThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMet 401
 Db 2407 AAAGAAAGATTAATGATTCATGCCACCTCCACCTCCACATTCCTCAATGCTCAAGATAG 2466
 QY 402 ThrThrThrValAsnTyrGlnAspGlyGlyLysValAlaValLeuCysLysGluAsnTyr 421
 Db 2467 ACAACTACAGTGAATTAATAGATGAGAGAAAAAATATCATTCCTGCTCAAGAAAACTAT 2526
 QY 422 LeuLeuProGlnAlaLysGluIleValCysLysAspGlyArgTrpIleSerLeuProArg 441
 Db 2527 ATATTTCAAGATGGAGAAAGAAATGTGTGCAAAGATGGAGATGGCAGTCCAGCCAGC 2586
 QY 442 CysValGlu----- 444
 Db 2587 TGTGTTGAATAATTCATGTTCTGACACCACTGAAATAGACCATGGAACATTCAGTCA 2646
 QY 444 ----- 444
 Db 2647 CCTGCTCTGCAGAGAAAGAGAGAGAAATGTTGAGCCCAAGAAATATGACATGCGAC 2706
 QY 444 ----- 444
 Db 2707 AGCTTAAGTTATCTTGTGAAGATGTTGCGAGATATCTGAAAGACATATATGCCAC 2766
 QY 444 ----- 444
 Db 2767 ATGGAAATGAGATCTTCTTGCCCTGAGTGTGGTGTGCTTGTGCACACACCTGAG 2826
 QY 444 ----- 444
 Db 2827 ATTCAAGACGTGTGTGCTGCCTCAGAGAAAGACACCTACCTGTACGAGAGAAAGTTACC 2886
 QY 444 ----- 444
 Db 2887 TACAGCTGTGATGAAGTTTGGATGTGATGACCTGCATCTATTAAGATGTTTAGAGAA 2946
 QY 444 ----- 444
 Db 2947 AATGCTCCCATCTCCAGAAATGCAATTAACAGAGATTTGTTTGACTTAACCCATTCAT 3006
 QY 444 ----- 444
 Db 3007 GACGCTGTAGTATAGGCCCAAGAGCCATGTAATAGTCAGAGAGACAGTGCATTTT 3066
 QY 444 ----- 444
 Db 3067 AATGTCTGCATATTTTCAATTTGGAAGACCTTAACACCATAGGTTGATTTTAGAGACAGA 3126


```

OY 444 ----- 444
Db 3127 TGGATAGGAAGCAGCATGCAGAGATGTTCTGTGTGATTCACCCAGTGTGGAAC 3186
OY 444 ----- 444
Db 3187 GCTGTATCTAGATAGAGGTCAAGGTATCAACCTGCTGAGAGAGACGTTATGAGTGC 3246
OY 444 ----- 444
Db 3247 ATCGAAGCTTACGCCCTTTATGGGATGAGAAGTGATGTTAATGGAAGTTGGACC 3306
OY 445 ----- 445
Db 3307 AANCCAGCTCAGTGCAGAACGCTTAAGGAANAATGTGGGCGACCCACCCTATAGACAT 3366
OY 458 GlySerPthThSerPheProLeuSerValTyrProProGlySerThyValThrTyrArg 477
Db 3367 GGAGACAGCAGCTCATTTCCCATTTACAGCTATCCCGGAACAGCTCGTACGATACCA 3426
OY 478 CysGlnSerPheTyrLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyr 497
Db 3427 TCCCACTCTACTATGAACTTGGAGGACAGCTCCATTAATGTAGAAATGACAGCTGG 3486
OY 498 SerGluProProArgCysLeuAspProCysValValSerGluGlnAsnMetAsnLysAsn 517
Db 3487 TCAGAACCCACCAAAATGCTTAGTGCATGTGTAGTTTCAGAAATGATGAGAAACAT 3546
OY 518 AsnIleGlnLeuLysTyrPArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaVal 537
Db 3547 AACATAGAGCTTAATAAGGACCATTAATAAACTTATTCAGAAACATGATACTATT 3606
OY 538 GluPheGlnCysLysPheProHISLysAlaMetIleSerSerProProPheArgAlaIle 557
Db 3607 GAATTTCCGTGTAGACAAAGATATTTATAGAGAGAGCCCACTACATATTCAGAGCAACC 3666
OY 558 CysGlnGlnLysPheGlnLysTyrProLysCys 568
Db 3667 TGTCAAGCAAGCAAACTGCTATCTACTACTGT 3699

RESULT 6
MUSCFHRD 2955 bp mRNA linear ROD 12-JUN-1993
LOCUS Mouse complement factor H-related protein mRNA, complete cds, clone
DEFINITION 9C4.
ACCESSION M29009.1 GI:192563
VERSION M29009.1 GI:192563
KEYWORDS complement factor.
SOURCE Mouse (strain C57/B10.WR) liver, clone 9C4.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2955)
Vik,D.P., Munoz-Canoves,P., Kozono,H., Martin,L.G., Tack,B.F. and
Chaplin,D.D.
Identification and sequence analysis of four complement factor
H-related transcripts in mouse liver
J. Biol. Chem. 265 (6), 3193-3201 (1990)
COMMENT Draft entry and computer-readable sequence for (J. Biol. Chem.
(1990) in press) kindly submitted
by D.P. Vik, 13-OCT-1989.

FEATURES
Source 1..2955 Location/Qualifiers
CDS
/organism="Mus musculus"
/db_xref="taxon:10090"
69..2495
/note="complement factor H-related protein"
/codon_start=1
/db_xref="GI:309166"

```

```

/translation="MGFCSMLLSNLTITAMLTSTAKGEVKSCEPFQFKRYLFEEL
RNPVPSIGNKSYKQDNQSPSGIFNMYLCTQVGMKRPVCRKQEHVNEEF
AYMKRIYVQGGSLKXQCYGVSGLONGODMTENMGSPKICRIKCSVDIEN
GFFSESPRTYALNRETSYRCQGYVTNTESTRITCLONGSPDSCIKSEBPPE
NSVTKNSSTPFEKLNLDLECLIGHENEFKHKRGSTTCTYGVMSDPSQYELESPV
LDRKLVSPPKREKRVGDLLEFCRSHGNEVPSQCYHFGSPSPKQGVASQ
PPEIDNGELNGARVEYSHGVDYDCRFELFKGNKLCQVDMATLTPVCEERT
CGDIDPELHGSYKFSVPYHGDSDVEFTCAETFTMGLGSVCLSGKWTQLPQVARD
OLEKCRVLTSTDLPAIKPKRNEFHNSITVYKCRDQKOEHSICINGKRNPNCTRK
TSCPPRPQIPLNLTIVETTYKTDGELSLVLCODNTITDPEMKCKDQWQSLPHIC
LPQGPSPILRGTSLLESLVQHEVYHCSGTFRIDGPAFKCEGKWSQSPKIC
TNCQDVLPTLEMAIRGKKRSYRKGQVYFROQSPQOMNGSDTVICVNSRWIKPKQ
DSRGKQPPPIPDNDGTSLSPVEYFSSVDQCYKYLKGTITTCRNKWSPP
TCLHACIPEINEMEARKIILKWRHFNISYHSGEDIEFCYKGYHARSGSPFRKCI
SGTINPQCE"
BASE COUNT 990 a 550 c 577 g 838 t
ORIGIN
Alignment Scores:
Pred. No.: 6, 41e-127 Length: 2955
Score: 1524.00 Matches: 292
Best Local Similarity: 47.02% Conservative: 87
Best Local Similarity: 36.23% Mismatches: 187
Query Match: 47.88% Indels: 240
DB: Gaps: 6
US-09-805-337A-2 (1-569) x MUSCFHRD (1-2955)
OY 1 MetLeuLeuPheSerValIleLeuIleSerTyrValSerThyValGlyGlyGly 20
Db ATGTGCTCTTATCCAAATATCCCTCACTGATGCTGCTTCCACTCTAAAGGGAAGTG 143
OY 21 ThrLeuCysAspPheProLysIleHisLysGluPheLeuTyrAspGluLysAspTyrAsn 40
Db AAATCTTGTAATTTCCAAATTCAAATATGACAGCTGTATTTTAAAGATCCCGACA 203
OY 41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheVal 60
Db CCCAAGCTTCCAGTATCTATAGAAATTAAGTACAGCTATAAGTGACAAAGGGGTTTCA 263
OY 61 SerProSerLysSerPheThrPheArgIleThrCysThrGluGlnLysTyrSerProThr 80
Db CCACCTTCTGGGCTTTCTGGGACTACTCTGTTGACAGTACAAAGGTTGGAAGCCCTGAA 323
OY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100
Db GTCCCATGTGTCAAGAAATGTGTTTCCATATGTGAGATGAGAAATTTGCATCTG 383
OY 101 GlyLeuIleHisLeuGluLysAspThrValGlnIleIleCysAsnThrGlyTyrSerLeu 120
Db GAAAAAATATATGTGACAGGCTGCTTTAAAGTCCAGTGTATATGCTATAGCTT 443
OY 121 GlnAsnAsnGlnLysAsnLysSerCysValGluArgGlyTyrSerThrProProLysCys 140
Db 444 CAAATATGTCAGACACATGATGATGACAGAAATGGCTGCTCCCTCTCCCAATATGC 503
OY 140 ----- 140
Db 504 ATCCGATATCAAGCATGTTTCAGTATGCAATATGAANTTGAAGATGGGTTTTTTCTGAA 563
OY 141 SerPhe----- 142
Db 564 TCTTTTCGTACATATGCTCTAATATAGAAACATCTCTATAGATGTAAACAGGATATGTC 623
OY 142 ----- 142
Db 624 ACAATATCTGGAGAAACGTCAAGATCAATAACTTGCTTCAAAATGATGATGCTCAAC 683
OY 142 ----- 142
Db 684 CCTCATGTATTAGTCTTGTGAAGAACCAAGTATTTGAGAAATTCGTAACTAAGAAATAT 743
OY 142 ----- 142

```

Db 744 AGTACATGCTTAAAGCTCAACGACAAATAGACTATGATGCTCTCATGACATGAAAT 803
 Qy 143 -----Thrlsgly----- 145
 Db 804 GAATATTAACATACCAAGGCTCTATACGCTACTTATTATGATGGTCTGATACGCC 863
 Qy 146 -----glucyshsvalproileleuglualaasnvalaspaagipro 160
 Db 864 TCATGTTATGAATAGATGACGCGTTCCCATTTCTAGACGAACACTGATCGTTCTCC 923
 Qy 161 Lyslysglyserlytyrlyvalglyaspvalleulyspheserlyarglyasnleuile 180
 Db 924 ACAGAAAGAAATACAGAGTGTGAGATTGTAGAAATCTCCGCCGCTCAGAGA---CAC 980
 Qy 181 Argvalglyseraspservalglncyslyrproglinphelgltprserproasnpherothr 200
 Db 981 AGAGTTGGGCCAGATTCACTGCTCAATGTTACCACTTTGGATGGTCTCTAGTTCCCTACA 1040
 Qy 201 Cyslysglygluvalargsercysglyproproproglinleuserasnlygluvallys 220
 Db 1041 TTAAGAGTCAAGTACATCATGTGCAACACCTCGTAATTCCTAATGGGAATTATAT 1100
 Qy 221 Gluilearglygluvaltyrlyrlyhsasngluvalgluttyrlyrlyaspysasnproasn 240
 Db 1101 GGAGCAAAAGAGTGAATACAGCCATGGTGGAGTGGATATGATTCGAACCTCAGAA 1160
 Qy 241 Pheileileasnlyprolyrlyrlyleuglycysvalaspglyglutprthrleupro 260
 Db 1161 TTCCTACTAAGAGGACCCAAATTAATCCAGTGTGTTGATGGATGGCAACCTTGCT 1220
 Qy 261 ThrCysvalgluvallysThrCysglytyrlyrlyleprogluvalgluttyrlyrlyval 280
 Db 1221 GTATGCTTGAGAGAGGACACATGTGGAGACATTCCTGAACATGACATGGCTCTGTC 1280
 Qy 281 GluProserValProProtyrlyrlyInhslyvalserValgluValasnCysArgasnly 300
 Db 1281 AGGTTTCTGCTCCCTCCCTACCAACCATCGAGATTCAGTGGAGTTCCTGTCGACAGAAC 1340
 Qy 301 TyrAlaMetileglyAsnAsnMetilethrcysileasnlyrlyleprthrleupro 320
 Db 1341 TTCACATGATGAGACTGGTGGTCACTTCTGCTTGAAGGAAAGGACCCAGCTTCT 1400
 Qy 321 MetCysvalAlaIleThHisInleuLysArgCysLysIleAlaGlyAlaSnIleLysThr 340
 Db 1401 AATGTGTTGCAACAGACCACTGGAGGTAGAGTGTGAGTCAAGTCAACTGACATAGCA 1460
 Qy 341 LeuLeuLysLeuSerGlyLysGluPheasnHIsAsnSerArgIleArgtyrArgCysSer 360
 Db 1461 GCAATTTAAACCAAAAAGAAATCAATTTTCAGCACTACCTCACCATGATTTACAAATGTAGA 1520
 Qy 361 AspIlePheArgTyrArgHisSerValCysIleasnlyLysTrpAsnProgluValAsp 380
 Db 1521 GACAGACAGAGTATGAAACATCAATCTGATCAATGAAATGGAATCTCGAACCAAC 1580
 Qy 381 CysThnGluLysArgGluInPheCysProProProProglinIleProAsnAlaGluAsn 400
 Db 1581 TGTAAAGGAAACATCC-----TGCCCTCTCCACCAAGATTCCAATTCACCTAGTG 1634
 Qy 401 MetThrThrThValAsnTyrglnAspglygluLysValAlaValleuLysLysGluAsn 420
 Db 1635 ATTTGAACACACTGTGAAATCTTGGATGAGAAAATTTATCTGTTCTTTGCCAAGACANT 1694
 Qy 421 TyrLeuLeuProgluAlaLysGluIleValCysLysAspglyArgTrpInSerleupro 440
 Db 1695 TACCTTAACACAGACCCAGAGAAATGATGTCAAAAGATGGAAGTGGCAGTCATTACT 1754
 Qy 441 ArgCysVal----- 443
 Db 1755 CACTGCATTGAGACTCTTGAGACCTCCACCTCAATTCCTGCTGACTGTTCTCTCT 1814
 Qy 443 ----- 443

Db 1815 GAGCTAGAGAGTTACCAACATGGGGAGAGGTTACATACCATTTGTTCTACAGCTTTGGA 1874
 Qy 443 ----- 443
 Db 1875 ATTGATGACACAGCATTTATTAAATGTGAAGAGAAAGTGGTGTGACCCACCAAAATGC 1934
 Qy 443 ----- 443
 Db 1935 ATAAAAAGCAATTGTGACGTTTACCACCAATTGAAAAATGCCATATAGAGAAAGAG 1994
 Qy 443 ----- 443
 Db 1995 AAAAAATCATATAGACAGAGAGAGAACAGTGCATTGATGCAATCTCTTATCAAAATG 2054
 Qy 443 ----- 443
 Db 2055 AATGGCTCAGACACTGTGACATGTGTAAATACCGGTGATTTGGAAGCCATATGCAAA 2114
 Qy 444 GluSerThrAlaTyrcysglyproproproserileasnAsnlyAspThrThrSerPhe 463
 Db 2115 GACTCAAGAGGAAATGTGGGCTCCCTCCACCTATTGACAAATGAGACATCACCTCTTG 2174
 Qy 464 ProLeuSerValTyProProglYserThrValThrTyArgCysGlnSerPheTyrls 483
 Db 2175 TCATTACAGCAATATGAAACATTCATCATGACTATACAGTGCAGAAATATATCTC 2234
 Qy 484 LeuGlnGlySerValThrValThrCysArgAsnLysGlyTrpSerGluProProArgCys 503
 Db 2235 CTTAAGGAGAAACAGCAATATACATGATGACAAATGCAAACTGCTGAGCCACCAACATGC 2294
 Qy 504 LeuAspProCysValAlaSerGluGluAsnMetAsnLysAsnAsnIleGluLeuLysTrp 523
 Db 2295 TTACATGCACTGTATATACCAAGAAACATTAAGAGACAGCAAAATATATTCTTAAATGG 2354
 Qy 524 ArgAsnAspglyLysLeuAlaIleThrcysArgAsnLysThrGlyAspAlaValGluPheGlnLysPhe 543
 Db 2355 ACACACACAGAAATATTTATTTCCATTCAGGAGGAGATTTGAATTTGAATGTAATAT 2414
 Qy 544 ProHisLysAlaMetileSerSerProProPheArgAlaIleCysGlnGluLysPhe 563
 Db 2415 GGATATCATAAAGCAAGAGATCACCGCATTTCTGTACAAAGTGCATTAGTGCACCATC 2474
 Qy 564 GluTyProIleCysGlu 569
 Db 2475 AATTATCCCATTTGTGA 2492
 RESULT 7
 RNO320522 4256 bp mRNA linear ROD 04-SEP-2001
 LOCUS
 DEFINITION
 Rattus norvegicus mRNA for complement inhibitory factor H (FH gene).
 ACCESSION
 AJ320522 GI:15485712
 VERSION
 complement inhibitory factor H, FH gene.
 KEYWORDS
 Norway rat.
 SOURCE
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 REFERENCE
 1 (bases 1 to 4256)
 Demberg, T., Goetze, O. and Schiاف, G.
 Rat complement factor H: molecular cloning, sequencing and expression in tissues and isolated cells
 JOURNAL
 Unpublished
 2 (bases 1 to 4256)
 Demberg, T.
 Direct Submission
 Submitted (07-AUG-2001) Demberg T., Immunology, Georg August University of Goettingen, Kreuzbergstr 57, 37073 Goettingen, GERMANY
 FEATURES
 location/Qualifiers
 1..4256
 /organism="Rattus norvegicus"

```

/strain="Sprague-Dawley"
/db.xref="taxon:10116"
/sex="male"
/risue_type="liver"
/clonelib="lambda Triplex"
/dev.stage="ages 10 - 12 weeks"
/risue_lib="Clontech rat liver 5' stretch plus cdna library"
24. .4209
/gene="FH"
24. .3734
CDS

```

```

/codon_start=1
/protein_id="complement_inhibitory_factor_H"
/db_xref="GI:15485713"
/translation="MRLSARIIMLILMTWCAEDCKGPPENSEISLGSMSSEOLYSE
GQAYQKRCPPGRTGLIVKCKNGKNGEWMPNSITCKRCKPGHGDMPGSLRLVAGS
EPEEAKKVVYCCDECYOLGIELIDRECDADMTNITLCEYKRLPTLELNGELIVSG
AAEPQOETVQCVARFECNSGFKIEGCKEMHCNSNGLSMNRKPCVCEISCLPPVENG
DGITLPIKRIENEROYKCKQGFVYKREGDVACGSGMNPDPSEMIETCLTPYIPNII
YTPHTRIKHIDDELRICEKNLIPATRSVPKCTITGMIPARSLCPDPQPKHMRH
LYEESRRPYPVPVIGKEYEYCYCNGGFTPTSPWMDYLRCTVNGMEEPVCLRCSV
LYEESRLYLMORYIEGOSAKVQCHSGSLPNDGDTILCPNGMSPPKVKRIKTSYV
SDIELENFSESDYTYLANKRTIRKCGQGVYNTGELISGITTLQDGMSRPSCILSS
CDMEPFEANMKRNNNTKRLNDKLDYEDCHIGLENEYKHKTKSITCIDGWSVPSCEY
RCEMPLIABHODLVYPPREVKRVGDSLFSFQRSGHVRGADLVQCYHFGMSNPFTECGE
OVKSCDQGLDIPNGEIKCTKRKEYSHGQVYEDCKPRFLKGPRLICVQBRKWTLLPI
CYEYRRTGCGDLPELEHSGSVKLSPVRYHHGDSVEFTGCEFTTGMIGHAVFCISGRWTL
POCVATDDELECKAKRSTGIDAIPRNKRKEFNHNNSVSTRCKOKOYEHSITCINRMDP
EPNCTRANKRCKPPEPPQILPNAQVLETTYKVLIDGKRVASLSDYEDHTTYSYVQDDOFLR
WOSLPRCEKIPCSQPPKITEHGSITKSPSSSEERDLIESSYEDHTTYSYVQDDOFLR
ENRTPICMNGKWSLSPRCVGIIPCGPPSPILPAIGISHLEESYQVEEYTVLMSSEFGIT
GPAFLKCYGCGMSSEPPKCTIKTDMDLPPFEELAKPPEKKRSKSEVEOVTFCPPPYEM
DGSIDTGVNTKMGIOAPVCNCSNCPNRPYNNATILPNRHKTKYSGDGRVADCKPPEE
LPMGSGSVPTGFTWTEPPKCKDSTGCKGPPRIDNEDHITLRSIPVATLSSVYQCC
NITLYLGNKLVYTCRNGKWSOPPTCLHAACVEDIDEMKHIVLYLRENAKIVYSGSENI
EFMCKPGKRGKSGPPPTKCIEGHINPTCV"
1194..4199

```

polyA_signal	4194. .419
polyA_site	/gene="FH" 4209

	1423 a	788 c	892 g	1153 t
BASE COUNT				
ORIGIN				

Alignment Scores:

Pred. No.:	2,846,121	Length:	4256
Score:	1463.50	Matches:	289
Percent Similarity:	40.90%	Conservative:	84%
Best Local Similarity:	31.69%	Mismatches:	172
Query Match:	45.98%	Indels:	367
DB:	10	Gaps:	4

US-09-805-337A-2 (1-569) x RNO320522 (1-4256)

QY 23 CysaspPheProLysIleHisHISgLyPheLeuTrpAspGluGuaAspTyrAsnProPhe 42
| | | | | : | | | | | : | | | | |
Db 996 TGGATTTTCACACATTCAACAATGAGAGCTCTGTATTATGAAGAAGCCCGAGACCCCTAAC 1055

QY 43 SerGlnValAlaProThrGlyGluValaPheTyrTyrSerCysGluTyrAsnAphaValSerPro 62
 1056 TTCCAGACCTATAGAGAAAGAGTACTATTACTTGACACAGGGGTATTACAGCGCT 115
 Db

Cy		63 SerLysSerPheThrProThraGlleThrcystHcgluglGlyTrpSerProthrProlYs	82
Dd	1116 TCACAGTCATCTACTCGCACTACCCTTGTCGCACACAGTAATAATGGGTGGAGCCTGAAGTTCCA	1175 ::: ::: ::: :::	

QY 83 CysLeuAqMecCysSerPheProPheValLysAsnGlyHisSerGluSerGlyLeu 102
 1176 TGCCCTGAGCAGCAATGATATTTCCATTATGCGAATATGCGAATCTTTTACTGCAAGA 1235
 Db

```

QY      103  ILeHLSLeuGlUGlGyASpRThValGlnIleIleCysAsnThrcIyTYSerLeuGlnAsn 122
          :::::|||||  ::  :::::  |||:::|||||  |||

```

Db 1236 AGATATATAGAGGTCAGTCTGCCAAAGTCAGTGTCAACAGTGGCTATATAGCTTCCAAAT 1295

QY	122	AasmilwysanrlseSerCysValgluarglTyrPsethThrproIleCys-----	140
DbB	1296	GGTCAACATACAATTATGTAAGACAGAATAATGGCTGGTCCCTCTCCCAAATCGTCGCCT	1355

QY 140 ----- 140

DB	1356	ATCAGACATTGTCAGTATCCAGATATAGCAATTCGAT	1415
QY	140	-----	140

Db 1416 TATACATATGCTCTAAATAGAAAAACACCGTATGATGTAAACAGGATATGTACAAT 1475

1476	ACCGGAGAAATATCAGGAATATTACTTCTCTTCAGATGGATGGTCACCTCGACCCCTCA	1535
140	-----	140
0y	-----	

[illegible]

DB	1338	19CAG11AAGCCTTGTGATGATGCTTGTATTTGAGAGATGCTATGAC1AAGAGATAATATACACA	1339
OY	140	-----	140

Db	1596	TGGTTAACTCAATGACAAATTAGACTATGAATGTCACATTGAGATATGAAAAATGAATAT	1655
Ov	1441	Seppethbctusgiv-----	1445

[illegible]

QY 146 -----GlucCysHisValProIleLeuGlulnaIaasnValaspAlaGlnProLysLys 162
1716 TATGAAGAGATGCGAGCATTCCTCCCTGTTACACCAAGACATTAGTGGTTTTCCAGAGAA 1775

QY 163 GlnSerTyrLeuLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIleArgVal 182

Db 1776 GTAAATACAACTGGACATTGGTAGTTTCCTCTGGCCGTCACGA---CACAGAGTT 1832

Db 1833 GGAGCAGATTTAGTGCATGCTACCACATTGGATGTCCTCCCTAATTTCCCAACGTGTAA 1892

Qy 203 GYGVNValArgSerCysGlyProProProGlnIleuSerAraGlyGluValLysGluIle 222
| | | | | : | | | | | | | | | | | | | | | |
Db 1893 GGCCAGTAAATCATGTGACCAACCTCTGAATCCGCATGGCGGAATAAAGGAACA 1952

QY 223 ArgLysGluGlnLutyrClYrHisAsnGluValaGluLutyrAspCysAsnProAsnPhelle 242
:::||||| ||||| ||| :::::::::::::::::::::::::::: |||| |:::|:::

Db 1953 AAAAAGTGAATACAGCCATGCTGACGCTGGTGAATATGATTTCGCAACCTAGATTTCCTA 2012

Db 2013 CTGAGGCGACCCATTAATCCAGTGTGTGACGGGAGATGGACACCTTGGCCATATGC 2072

Oy	263	VaIsgInGInValIyThrcysGILTyrrIleproGInIeugILyrgILyValgInpro	282
		:::	:::
Db	2073	GTTGAGTATGAGACACATGTGGAGACCTTCCTTAACATTGAGCATGGCTCTGTCAAGTTA	2132

QY 283 SerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGluTyrAla 302
|||||
|||||

DB 2133 TCTGTCCTCCCTACCAATCATGGAGATTTCAGTGGAGTTTCACTTGTATACAGAAACCTTCACA 2192

QY 303 MetIleGlyAsnAsnMetIleThCysIleAsnGlyIleTrpHrgIleuProMetCys 322

Db 2193 ATGATGGACATCAGTACTTTTCTGCATTAGTAGGAAGGTGGACCGAGCTTCCCAATGT 2252

	325	valatdihrlstglnleulytsrlysllealadglyvlaasnlleyssirleuen	342
QY			:::
	2253	gttgcacagcatcactggagaactgttaaaccccggaactcaactggcattagatgcatt	2312
Db			:::

Oy 343 LysLeuSerCylLysCluPheAsnHisAsnSerArgIleArgTyrArgCysSerAspIle 362
 :: ||||| :: |||||
Db 2313 CATCCCAATTCAGATCTATTAACTTTACCTGCGGGAGCTGCATGCTCATGCCAAGA 2372

OY	363	theatgTATATGHSLeSeValCysTLeasnGlyLysTrrPanProGluValAspCysThr	382
Db	2373	CAGGAGTATGACATTCATCAATCGCATGGAATGGATCTCGAACCACAGCTACA	2432
OY	383	GlutAsrGluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMetThr	402
Db	2433	AGAAATGAGAAATGATCTGCCCTCCCTCCCTCCACAGATTCCAAATGCCCAAGTGATGAA	2492
OY	433	ThrThrValAsnTyrGlnAspGlyGluLysValAlaValIleCysLysGluAsnTyrLeu	422
Db	2493	ACCAAGTGAATACTGTGATGAGAGAAATATCTGTTCTTGGCAAGATGGTTACCTA	2552
OY	423	LeuProGluAlaLysLysLysIleValCysLysAspGlyAsrGTrPglSerLeuProArgCys	442
Db	2553	ACTCAGGGCCCAAGAAATAGTGtGTGTAACATGGAAGCTGGCACTGTTACCCAGCTGC	2612
OY	443	ValGlu-----	444
Db	2613	ACGAAAAAATTCATGATGTTCCAGCCCCCTAAATATGACATGATCTATTAAAGTCGCC	2672
OY	444	-----	444
Db	2673	AGGTCCCTAGAAAGACAGAGATTAAATTGATGCCAGCACTTATGAACCGAACAATTC	2732
OY	444	-----	444
Db	2733	AGCTATGCTGTGATGATGATTCAGATATCTGAAGAATAAGGGTAACTCGAACATG	2792
OY	444	-----	444
Db	2793	GGAAATGAGACTCTCTGCCCTGCTGTGTGGAATACCTTGAGACCCCACTTCATTT	2852
OY	444	-----	444
Db	2853	CCTGTGTGATTGTTTCTCATGAACTAGAAATTACCAATATGAGAGAGATTACATAC	2912
OY	444	-----	444
Db	2913	AATTGTTCTGAAGGCTTGTGAATTGAGGACGACGATTTATTAAATGTTAGAGAGACG	2972
OY	444	-----	444
Db	2973	TGCTGTGAACCTCCCAAAATGCAATAAAACTGATTGTGCAACTTGGCCACATTTGAATT	3032
OY	444	-----	444
Db	3033	GCCAAACCGACAGAAAGAAAAAATCATATACAGTCAGAGAAACAAGTGCATTCGAA	3092
OY	444	-----	444
Db	3093	TGTCACCTCCGTTATCGAATGATGCGCTGTACATTTGTACATGTGTTAAATACAAGTGG	3152
OY	444	-----	444
Db	3153	ATTGGACAGCGGATATGCAAAATTAATCTGTGTGAATCCACACATGTGCCAAATGCT	3212
OY	444	-----	444
Db	3213	ACTATACTAACAGGACACAGCATTAATATTCATCTGTGACAAAGTACATTAGACTGT	3272
OY	444	-----	444
Db	3273	AATAAACCCTTTGAATTATTTTGGGGAAGTGAAGTCAATGTGCCCCAAACGGGATTTTGG	3332
OY	445	-----	445
Db	3333	ACAGAACCCAGCAATGCAAGATTCAACAGAGGAAATGTGGGCTCTCCACGATATGAC	3392
OY	457	AsnGlyAspThrThrSerPheProLeuSerValTyrProProGlySerThrValThrTyr	476
Db	3393	AATGGAGACATCACCTCCTGTGTCAATTAACAGTATATGACCATTAATCATCATGTTGAATAT	3452

OY		477	Argcysglnserpheptyllyslengluiglyceralthrvalthrpcysargasnllysgln	496
Dd		3453	Caahtgccgaacttattaccttaagggaataatagatgcattgacaaaatcgcaaac	3512
OY		497	trpserglupropiorargcysleuaasperocysvalValSerugluAsnmetasllys	516
Dd		3513	TGTCGTCCAGCACAACCTTGTTACATGCATGTGGATTACCAGAAGATTATTATGAANA	3572
OY		517	Asnsanllleglnleullystprargasaapdllylsleutyralalystrhglyaspala	536
Dd		3573	CATAATATAGTCTCGACATGGAGGAAATGCCAAGATTATATCCCAATCCGGGACAAAT	3672
OY		537	VAlcluphegnlcylsypheprohlisylsalamelilesersepproPheargla	556
Dd		3633	ATFGAATTCAACTGTGAACCAACGTGATATAGAAAATTCAGAGAGCATCACCTTCGTGA	3692
OY		557	IlecysglingluiglylyspheglutyrProilecys	568
Dd		3693	AAGTGCATTTGAGGGTCACATCATTAATTCCTCCCTTGT	3728
RESULT 8				
LOCUS	A91883	4229 bp	DNA	linear PAT 22-JAN-2000
DEFINITION	Sequence 1 from Patent WO9823638.			
ACCESSION	A91883			
VERSION	A91883.1	G1.6740760		
KEYWORDS	unidentified. unclassified.			
SOURCE ORGANISM	unclassified. unclassified.			
REFERENCE	1 (bases 1 to 4229) Schwaebtle,W. and Stim,R.B. COMPLEMENT INHIBITOR Patent: WO 9823638-A 1 04-JUN-1998; SCHWAEBLE WILHELM (GB); UNIV LEICESTER (GB)			
AUTHORS TITLE JOURNAL FEATURES	Location/Qualifiers source 1..4229 /organism='unidentified' /db_xref='taxon:32644'			
BASE COUNT ORIGIN	BASE COUNT 1395 a 791 c 893 g 1150 t			
Alignment Scores:	Pred. No.: 1.48e-120 Length: 4229 Score: 1455.50 Matches: 289 Percent Similarity: 40.7% Conservative: 83 Best Local Similarity: 31.6% Mismatches: 171 Query Match: 45.73% Indels: 369 DB: Gaps: 5			
US-09-805-337A-2 (1-569) x A91883 (1-4229)				
OY	23 Cysaspheprolysllehshishglyphelautyraspgluasptyrasnrophe	42		
Dd:: ...	1064		
OY	43 SerqlnalprothrngluValphertytyrsercysgluyrasnphevalserpe	62		
Dd	1124		
OY	63 serlysserpherptrhrgllerthrCysthringluigltryrsetPrpthrProlys	82		
Dd	1184		
OY	83 CyslleargmetycysserpheprophevallylasncllyhisserciuserSergyleu	102		
Dd	1244		
OY	103 llehistengluiyaspthrvaGlleiellcysansnthrglytyrserIeuGlinaSn	122		
Dd	1304		
OY	123 AsnglulysasnlliesercysvalgluaragltryrsetThrproplecys-----	140		

[illegible]

Db	2382	CAGGAGTATGAAACATTCAATCATGTCATGCATGGAAGATGGATCTCTGAACCAAACTGTACA	2441
QY	383	GIuYsArGluGIuGlnPheCysProProProGlnIleProAsnAlaGlnAsnMetThr	402
Db	2442	AGCAAAAGA-----TTCTGCCCTCTCTCCCGGAGATTCCAATGGCCCAAGTATTGAA	2495
QY	403	ThrThrValAsnTYrGlnAspGlyGluYsValAlaValIleuYsLysGluAsnTYrIleu	422
Db	2496	ACCAACCGTGAATACTTTGGATGGAGGAAAAAGATATCTGTCTTGGCCAAGATCGTTACTTA	2555
QY	423	IeuProGlnAlaLysGluIleValIcysLysAspGlyArGTPpGlnserIeuProArGys	442
Db	2556	ACTCAGGCGCCAGAAAGAAATGCTGTAAACATGGAAAGGTGGACGTCTGTACCAACGCTGC	2615
QY	443	ValGIu-----	444
Db	2616	ACGGAAAAAATTCCATGTTCCACGCCCCCTAAATTGAACATGATCTATTAACTCGCC	2675
QY	444	-----	444
Db	2676	AGTTCCTCAGAAGAGAGAGATTAAATGAGTCACAGACTTATGAACACGAACTACA	2735
QY	444	-----	444
Db	2736	TTTCAGCTATTGCTGTAGAGATGATTCAGATATCTGAAGAAATAGGGTAACCTGCAAC	2795
QY	444	-----	444
Db	2796	ATGGGAATATGAGAGCTCTCTGCTGCTGTGTGTGTGAATACCTTGTGGACCCCACTTCA	2855
QY	444	-----	444
Db	2856	ATTCCCTTGTGATTGTTTTCATGACACTAGAAGATTACCAATATGAGAGAGGTTACA	2915
QY	444	-----	444
Db	2916	TACAAATGTTCTGAAGGCTTTGGAAATGATGAGACAGCAGCTTATTAATGTGTAGAGAGA	2975
QY	444	-----	444
Db	2976	CAGTGCCTGAACCTCCCAATGCAATAAACTGATTTGACAACTGCCCACTTTGAA	3035
QY	444	-----	444
Db	3036	ATTGCCAAACGACAGAAAGAAAAAATCATACAGTCGACGAGCAAGCAAGTGACATTTC	3095
QY	444	-----	444
Db	3096	AGATGTCACCTCCGTATGAAATGATGCTGTGACATTGTACATGTGTTAATACGAAG	3155
QY	444	-----	444
Db	3156	TGGATTGACACCGCGTATGCAAGATATTCCTGTGTGAATCCACCACTGTGCCAAAT	3215
QY	444	-----	444
Db	3216	GCTACTATACTAACAAGGCACAAGACTAAATATCCATGTGGTGACAAAGTACGTTATGAC	3275
QY	444	-----	444
Db	3276	TGTAATAACCTTTGAATTATTTGGGAGAGTGAAGTATGTCAAAGCGGATTTGG	3335
QY	445	-----SerThrAlaTYrCysGlyProProProSerIleAsn	456
Db	3336	ACAGAACCAACGAAATGCAAAAGATTCAACAGGGAATGTGGGCTCTCCACCTATTTGAC	3395
QY	457	AsnGIAspThrThrSerPheProIleuSerValTYrProProGlySerThrValThrTYr	476
Db	3396	AATGAGACATCACCTCTTGTCAATTACACGATATATGACACATTATCATCTCAGTTGAATAT	3455
QY	477	ArGysGlnSerPheTYrLysLeuGlnIYserValThrValTYrCysArGAsnLysGln	496
Db	3456	CAATGCCGAACATATTATCTACTTAAAGGAATAGATAGTACATGTAGAAATGGAAG	3515

```

OY 497 TrpSerGluProArgCysLeuAspProCysValIleSerGluGlnAsnMetAsnLys 516
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3516 TCGTCTACGCCACCACTGCTTACATGATCTGCATACCAAGATATTGTGAAAA 3575
OY 517 AsnAsnIleGlnLeuLysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAla 536
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3576 CATATATATAGTCTCAGATGAGGAAATCAAGATTATTCCTCAATCAGGGAGCAT 3635
OY 537 ValGluPheGlnCysLysPheProHisLysAlaMetLiesSerProPheArgAla 556
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3636 ATTGAATTCATCTGTAACTGCGATATAGAAATTCACAGCATCACTCCGTTGTGACA 3695
OY 557 IleCysGlnGluGlyLysPheGluTyrProIleCys 568
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3696 AAGTCATTCAGCGTCACATCAATATATCCCACTTCT 3731

RESULT 9
AF436847 2729 bp. mRNA linear ROD 23-APR-2002
LOCUS Rattus norvegicus complement factor H-related protein mRNA,
DEFINITION complete cds.
ACCESSION AF436847
VERSION AF436847.2 GI:20271386
KEYWORDS
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
          Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
          Eukaryota; Metazoa; Rodentia; Sclurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 2729)
AUTHORS Ren, G. and Quijg, R.J.
TITLE Rat complement factor H-related protein sequence
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2729)
AUTHORS Ren, G. and Quijg, R.J.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2001) Nephrology, University of Chicago, 5841 S.
          Maryland Ave., Chicago, IL 60637, USA
REFERENCE 3 (bases 1 to 2729)
AUTHORS Ren, G. and Quijg, R.J.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Nephrology, University of Chicago, 5841 S.
          Maryland Ave., Chicago, IL 60637, USA
REMARK Sequence update by submitter
COMMENT On Apr 23, 2002 this sequence version replaced gi:16551387.
FEATURES
          Location/Qualifiers
            source
              organism="Rattus norvegicus"
              strain="Sprague-Dawley"
              db_xref="taxon:10116"
              cell_type="glomerular epithelial cells"
              tissue_type="kidney; also found in liver"
              codon_start=1
            product="complement factor H-related protein"
            protein_id="AAU25802.2"
            db_xref="gi:20271387"
            translation="MRFSIMLANILITSLSTAKGECEVLHVENGYSFLERK
              VODSVKVCYKGFENLPNGODTICTESGMDPOPICISFSCSDIDINDFSEED
              LTYAVNRTRYKGGYTTANGERSGISTCOTGMSAPSCDIDVDVAFENAKSKND
              STWFLNKLDECHGYENKXKHSIMKTOGMSKPSCEIEGIRPLDRLVY
              YPRKIKYVGLDKRSRGRHVRGPDVOCDFGMSPSPTCKGVRSGCGQPELLNG
              EIKGTAKKEYSHGDVVEYDCKRPLFLKPNKIQCYDGMWTLPTCEERKRTGDLREL
              EHGSKFSPYPHHGDVSEFTCAEFYVIGGSGVSCISGRMAQLPRCVATQLEKCA
              PRLAIVGINKLKDGLNHNSSISYKCGSGVYSICINGRMDPEPTCTREKFCGP
              PPOLPNHAVIETVYKYLDEKVKYSCDGYLTGCEEMVCKHGRMOSIPCTAKRTCS
              QPKRDHSITLPLRSDAGDAIESSEHRTTSPSYVDDCFRISKENGVCYCHKMS
              SPKCVESTGKCGPPTIENGDTLSLPLVAPLSSVEYQCSFVKMLGSKRTCKNG
              EMEEPKCLACVPLEEIMRRONTIIPKTKNKIYIOGSDYVEFVCLAYOKAOGSPF
              FRTOCIDGHINYPICSK"
CDS
          BASE COUNT 916 a 502 c 561 g 750 t
          ORIGIN

```

```

Alignment Scores:
Pred. No.: 3 67e-118
Score: 1426.00
Percent Similarity: 52.57%
Best Local Similarity: 40.57%
Query Match: 44.80%
DB: 10 Gaps: 11

US-09-805-337a-2 (1-569) x AF436847 (1-2729)
OY 3 LeuLeuPheSerValIleLeuLieserTyrPheValSerThrValGlyGluGlyThrLeu 22
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 CTTTACGAAATATATCTCTATACATCACTTCTTACTGCTAAAGGAGAGAA--- 127
OY 23 CysAspPheProLysIleHisIleGlyPhe--LeuTyrAspGluGlnAspTyrAsnPro 41
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 128 TGTGTTTGCATATATGTAAGAAATGTTATCTTATTCCTGGAAGAAATTTGTACAA 187
OY 42 PheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheValSer 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 188 GATCAGTCTGTCAAGTC-----AACTGTATAAGGCTTCAACTT 229
OY 62 ProSerLysSerPheTyrThrArgIleThrCysThrGluGlyTyrSerProThrPro 81
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 CCAATATGGTCAG-----GATCAATATATATGTCACAGAGAGTGGCGACACCTCAACC 283
OY 82 LysCysLeu-----ArgMetCysSerPheProPheValLys-----AsnGlyHisSer 97
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 ATATCATTCCTCTCGACATGCTCAAGATCAATAGATATAGAAATGATATGATTTT 343
OY 98 GluSerSerGlyLeuIleHisLeuGlnLysPthrValGlnIleIleCysAsnThrGly 117
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 344 TCTGATTCGATTTAATCATATGCTTAATAGAAACACGATATGCTGTAACAGCGA 403
OY 118 TyrSerLeuGlnAsnAsnGlnLys-----AsnIleSerCysValGlnArgGlyTyrSer 135
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 404 TATACACAGCAAGAAATGAGAAAGATCGGATCAATCACTTGCTCAACACTGATGCT 463
OY 136 ThrProPheCys-----|||||:|||||:|||||:|||||:|||||:|||||: 140
Db 464 GCTCAACTTCTTGATTAAGCTCTTGATGCTGCTGTTGAGAAATGCTAAATCTAA 523
OY 140 ----- 140
Db 524 AATGATACACATGCTTAACTCAATGACAAATAGATTAATGATGTCACATTCGATAT 583
OY 141 -----SerPheThrLysGly----- 145
Db 584 GAAATTAATATTAAGCATACCAAGGCTCCATATGCTACTACTGATGATGCTGAT 643
OY 146 -----GluCysHisValProIleLeuGlnAlaAsnValAspAla 158
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 644 AAACCTCATGTTATGAATTAATGAATGACGATTCCTCATATTAGACCACGCTTATGTT 703
OY 159 GlnProLysLysGluSerTyrLysValAlaLysAspValLeuLysPheSerCysArgLysAsn 178
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 704 TATCCAGAGAAATTAATATATACAGTTCAGATGTTGTAATCTCTTGCCAGACACAGA 763
OY 179 LeuIleArgValIleSerAspSerValGlnCysTyrGlnPheGlyTyrSerProAsnPhe 198
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 764 ---CACAGAGTTGACCGCATTCAGTTCGCAATGCTATGACTTTGGATGGTCCCTAAGTTT 820
OY 199 ProThrCysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGlu 218
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 CCGACATGTAAGGTCAGCAAGTAAGATCATGCTGTAATTCGCAACTCAATGAGGGA 880
OY 219 ValLysGluIleArgLysGluGlyTyrGlyHisAsnGlnValGlnTyrAspCysAsn 238
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 881 ATAAGGGAACAGAAAGCTGAATGACAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 940
OY 239 ProAsnPheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrThr 258
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Intron 3426..3473
 /note="H cds intron"
 exon 3474..4300
 /number=2
 BASE COUNT 1401 a 814 c 902 g 1183 t
 554 bp upstream of XhoI site; chromosome 1.
 ORIGIN

Alignment Scores:
 Pred. No.: 2,59e-117 Length: 4300
 Score: 1419.50 Matches: 283
 Percent Similarity: 39.33% Conservative: 82
 Best Local Similarity: 30.50% Mismatches: 178
 Query Match: 44.60% Indels: 385
 DB: 10 Gaps: 5

US-09-805-337a-2 (1-569) x MUSH (1-4300)

QY	23	CysAspPheProLysIleHISglYpHeLeuTYrAspGluAspTYrAsnProPhe	42
DB	1073	TCGTAATTTCCACAAATTGAGACGCTGTATATGACAGAGAGCTGAGACCCAAC	1132
QY	43	SerGlnValProThrGlyGluValPheTYrSerCysGluTYrAsnProPheValSerPro	62
DB	1133	TTCACGATCTATAGCAAAATAGTACAGCTTAAGTGTACACGGGTTTTCACACCT	1192
QY	63	SerLYSerPheThrPheArgIleThrCysThIGluGluGlyTYrSerProThrProLys	82
DB	1193	TCTGGGTATTCCTGGGACTACCTTCGTGCACAGCACAGGGTGGAGCCTGAAGTCCA	1252
QY	83	CysLeuArgMetCysSerPheProPheValIysAsnGlyHISerGluSerGlyLeu	102
DB	1253	TGGGTAGCAAAAGCTTTTCCATTATGTGGAGAAATGAGACTCTGCATCTGGGAAAA	1312
QY	103	IleHISLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTYrSerLeuGlnAsn	122
DB	1313	GTTATGTGCGAGGTGAGCTTAAAGTCCAGCTGTACAGTGTATAGTCTTCAAAAT	1372
QY	123	AsnGluLysAsnIleSerCysValGluArgGlyTYrSerThrProProIleCys	140
DB	1373	GCTCAAGACACATGACATGTACAGAAATGGCTGCCCTCCCAATGCATCCGT	1432
QY	140	-----	140
DB	1433	ATCAGACATGTTCCACATCAGATATACATGACATGATTTCTTCTGAAATCTCT	1492
QY	140	-----	140
DB	1493	TCTATATGTCTCTAATATAGACAACATCTATAGATGTAAGCAGGATATGTACAAT	1552
QY	140	-----	140
DB	1553	ACTGAGAAATATCAGATCAATTAACCTTCAAAATGATGTACACCTCAACCTCA	1612
QY	140	-----	140
DB	1613	TGCATTAACTCTTGATATGCTGTATTGAGAAATTTATATACTAAGAACTAGACA	1672
QY	140	-----	140
DB	1673	TGCTTAAGCTCAATGACAAATTAGACTATGAATGTCGTGGATTGAAATGAAATAT	1732
QY	141	SerPheThrLYsgly-----	145
DB	1733	AAACATACCAAGGCTCTAATACATGACTTATATGATGTCGTATACACCTCATGT	1792
QY	146	-----GluCysHisValProIleLeuGluAlaAsnValAspIleGlnProLys	162
DB	1793	TATGAAGGAAAGCGCTTCCACTAGACCCGAAACCTAGTCCCTTCCCCAGAAAA	1852
QY	163	GluSerTYrLYsValGlyAspValLeuLysPheSerCysArgLYsAsnLeuIleArgVal	182
DB	1853	GAATAATACAGACTTGAGATTGTTGGAAATCTCTGCCATTACAGCA---CACAGAGTT	1909

QY	183	GlySerAspSerValGlnCysTYrGlnPheGlyTYrPheSerProAsnProPheThrCys	202
DB	1910	GGGCCAGATTTCAGTGCATGCAATGCTACACCTTTGGATGTCCTCGGTTTCCCTCATGTAA	1969
QY	203	GlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyValLYsglyLeu	222
DB	1970	GTCAGATGATCATGTCACACCTCTGAAATCTTAATGGGAAATTAATGAGACA	2029
QY	223	ArgLYsGluGlyTYrGlyHisAsnGluValGluTYrAspCysAsnProAsnProIle	242
DB	2030	AAAAAGTTGAATACAGCCATGCTGAGTGTGAATATGATGCAAACTGATTCCTTA	2089
QY	243	IleAsnGlyProLysIleGlnCysValAspGlyGluTYrPheThrLeuProThrCys	262
DB	2090	CTGAAGGAGACCAATTAATTCAGTGTGTGATGGGAATTTGACAACTTGCTCATGT	2149
QY	263	ValGluGlnValLYsThrCysGlyTYrIleProGluLeuGlyTYrValGlnPro	282
DB	2150	ATTGAGAGAGAGAAACATGTGAGACATTCCTGAACTTGAACTGCTCCCAAGTGT	2209
QY	283	SerValProProTYrGlnHisGlyValSerValGlnValAsnCysArgAsnGluTYrAla	302
DB	2210	TCTGTTCTCCTACACCATGAGATTCAGTGTGAGTTCATTTGTGAACAAACTTCACA	2269
QY	303	MetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTYrPheGluLeuProMetCys	322
DB	2270	ATGATGTGACATGGCTCAGTCTTCTGATAGGAAATGAGACCACTGCTCTTAATGT	2329
QY	323	ValAlaThrHisGlnLeuLysArgCysLYsIleAlaGlyValAsnIleLYsThrLeuLeu	342
DB	2330	GTTGCAACAGCCAACTGGAGAAAGTGTAGAGTCTGAAATGCATACCTGCATAGACCAATA	2389
QY	343	LysLeuSerGlyLYsGluPheAsnHisAsnSerArgIleArgTYrArgCysSerAspIle	362
DB	2390	AAACCAAAATTCAGTCAATTTACGCAATCTACCATCATGATGATTAAGATGAGACAG	2449
QY	363	PheArgTYrArgHisSerValCysIleAsnGlyLYsTYrAsnProGluValAspCysThr	382
DB	2450	CAGGACTACGAAACGCTCAATCTGTATCAATGGAATGGATGCATCAACCAACTGTACA	2509
QY	383	GluLYsArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMetThr	402
DB	2510	AGCAAAACATCC-----TGCCCTCCTCCACGCGAGATTCCAAATACCAAGTATGAA	2563
QY	403	ThrThrValAsnTYrGlnAspGlyGluLYsValAlaValLeuCysLYsGluAsnTYrLeu	422
DB	2564	ACCACCGTGAATPACTTGATGATGAGAAATATATCTTCTTGGCAACACATTAACCTA	2623
QY	423	LeuProGluAlaLYsGluIleValCysLYsAspGlyArgTYrPheGlnSerLeuProArgCys	442
DB	2624	ACTCAGGACTCAGAAAGAAATGCTGTCCAAGATGGAAGGTGGCAGTACATTACCTGCTGC	2683
QY	443	ValGlu-----	444
DB	2684	ATTGAAAAAATTCATGTTCCAGCCCTACAAATAGAACATGATCTATTATTACC	2743
QY	444	-----	444
DB	2744	AGATCTTCAGAGAAAGAGAGATTCCATTGAGTCCAGCAGTCATGAACTGAACTACA	2803
QY	444	-----	444
DB	2804	TTACAGTATGCTGTGATGATGTTTCAGAGATACCTGAAGAAATAGATTAACCTGTAC	2863
QY	444	-----	444
DB	2864	ATGGGAAATGAGACACTCCACCTGCTGTGTGACATTCTTGGAGACTCCACCTTCA	2923
QY	444	-----	444
DB	2924	ATTGCTCTGTGACTGTTCTCTTGAAGTACAGAGATTACCAACATGGGGAAGGTTACA	2983
QY	444	-----	444

Db 2984 TACCATTGTTCTACAGCGTTTGGAAATGATGACACGACATTTATATATGCGAAGAGGA 3043
 444 -----
 Db 3044 AAGTGTCTGACCCACCAAAATGCATTAAGAGGATTGTGACGTTTACCCACAGTTAA 3103
 444 -----
 Db 3104 AATGCCATAATAAGAGAAAGCAAAAATCATATAGGACAGAGACAGACAGTTC 3163
 444 -----
 Db 3164 AGATGCAATCTCCTTATCAAAATGAATGCTCAGACACTGTGACATGTATTAAGTCGG 3223
 444 -----
 Db 3224 TGCATTGGACAGCCAGTATGCAAAAGTAATTCCTGTGTGATCCACCACATGTGCCAAAT 3283
 444 -----
 Db 3284 GCTACTATAGTACAGAGCAAGAAATATATCTACATGTGTGACAGAGTACGTTATGA 3343
 444 -----
 Db 3344 TGTAAATAACCTTGGAACTATTGGGCAAGTGGAAGTATGTGAAAATGGGATATGG 3403
 444 -----
 Db 3404 ACAGAAAAACCAAGTCCGAGCTCTGTAAATTCGACTTGAGTCTCAACCTTCAAAATGTT 3463
 445 -----
 Db 3464 TTCTCCTTAGACTCAACAGGAAATGTGGGCTCCCTCCACCTATGCAATGAGACATC 3523
 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThyTyrArgCysGlnSer 480
 3524 ACCTCCTGTGATACAGTATGACATGTAATGCAATTCATCAATATCATATCATATCCAGAAC 3583
 481 PheTyrLeuGlnGlnSerValThyValThyCysArgGlnSerGlnPro 500
 3584 TATATCTCTTAAAGGAAAGAACATATACATGTAACAAATGAAAGTGGCTTGACCA 3643
 501 ProArgCysLeuAspProCysValValSerGlnGlnAsnMetAsnLysAsnLysGln 520
 3644 CCAACATGCTTACATGATGTATATACGAAACATATATGCAATCACAATATATAT 3703
 521 LeuLysTyrArgAsnAspLysLysLeuTyrAlaLysThyGlyAspAlaValGlnPheGln 540
 3704 CTCAAATGAGACACACTGAAAGATTTATTTCCATTCAGGAGGATATGGAATTTGGA 3763
 541 CysLysPheProHisLysAlaMetLysSerSerProPheArgAlaLysGlnGln 560
 3764 TGTAAATATGATATATATAAGCAAGATTCACCCGATTTGCTACAAAGTCAATTAAT 3823
 561 GlyLysPheGlnTyrProLys 568
 3824 GGCACCATCAATATCCCACTGT 3847

FEATURES SCHWAEBLE WILHELM (GB); UNIV LEICESTER (GB)
 Location/Qualifiers
 source 1..2715
 /organism="unidentified"
 /db_xref="taxon:3264"
 BASE COUNT 863 a 518 c 618 g 716 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,79e-107 Length: 2715
 Score: 1307.00 Matches: 252
 Percent Similarity: 47.91% Conservative: 69
 Best Local Similarity: 37.61% Mismatches: 157
 Query Match: 41.06% Indels: 192
 DB: 6 Gaps: 7
 US-09-805-337A-2 (1-569) x A91885 (1-2715)
 23 CysAspPheProLysIleHisGlyPheLeuTyrAspGlnGlnAspTyrAsnProPhe 42
 660 TGTGATTTTCCAAATTCAAACATGAGACGCTGTATATGAAGAAAGCGGAGACCTTAC 719
 43 SerGlnValProThrGlyGlnValPheTyrTyrSerCysGlnTyrAsnProPheValSerPro 62
 720 TTCCAGTACCTTATAGGAAAGAGTACAGCTATATACCTGACAAACGGGTTTACACGCT 779
 63 SerLysSerPheThrPheThrArgIleThrCysThrGlnGlnGlyTyrSerProThrProLys 82
 780 TCACAGTCACTACGGGACCTACCTCTTGCACAGTAAATGGGTGGAGCCTGAAGTTCCA 839
 83 CysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGlnSerGlyLeu 102
 840 TGCCCTCAGGCAAGTCAATTTTCCATATGATGGAATATGGAATCTTCATACCTGCAAGA 899
 103 IleHisLeuGlnGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGlnAsn 122
 900 AGATATATATAGAGGTGAGCTGACGTCGAAAGTCCAGCTGACAGTGGCTTATGCTTCCAAAT 959
 123 AsnGlnLysAsnLysSerCysValGlnArgGlyTyrSerThrProProLys 140
 960 GTTCAGATACATATATATTTTACAGAGAAATGGCTGCTCCCTCTCCCAATATCGTCCGT 1019
 140 -----
 1020 ATCAGACTGTGTCAGATACAGATATAGAAATTTGAAATGGTTTTCTGCAATCTGAT 1079
 140 -----
 1080 TATACATATGCTCTTAATATAGAAACACGCTATATGTAACAGGAGATATGTAACAAT 1139
 140 -----
 1140 ACCGAGAAATATACAGAAATATACTTGTCTTCAAGATGATGTGCACCTGCACCCCTCA 1199
 140 -----
 1200 TGCATTAGCTGTGATATGCTGTATTTGAGAAATTTATGACTAAGATAATATATACACA 1259
 140 -----
 1260 TGGTTAAACTCATAGCAAAATTAAGACTATGAATGTGCATTTGGATATGAAATGAATAT 1319
 141 SerPheThrLysGly-----
 1320 AAACATACCAAGAGCTTATACATGATCTTATGATGAGATGCTGTAGTACACCTCTCTGT 1379
 146 -----GlyCysHisValProIleLeuGlnAlaAsnValAspAlaGlnProLysLys 162
 1380 TATGAAAGAAATGACAGCATTCCTCTGTATACACAGACTTATGTTTCCACAGAGA 1439
 163 GluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIleArgVal 182
 1440 GTAAATATACAAAGTTGAGATTCCTTGAGATTCTCTTCCGTCACAGCA---CACAGAGTT 1496


```

OY 201 CysIsglyValArgSerCysGlyProProGluLeuSerAsnGlyValValys 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 146 TGTAAGAGCAAGTACATCATGTGCTCACCCTGACCTCCATCCATGGAATGTTAAG 205
OY 221 GluIleArgLysGluGluTyrGlyHisAsnGluValValGluTyrAspCysAsnProAsn 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 GAAAAAGCAAGAAAGATATGACACAGTGAAGTGGAAATATATATGCAATCCCTGA 265
OY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrLeuPro 260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 TTCTCAAGGAAGGACCTTAATTAATTCATGCTTGAAGAGTGGACCACTTACCA 325
OY 261 ThrCysIleGluValIleLysThrCysGlyTyrIleProGluLeuGlyTyrGlyVal 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 326 GTGTATATGAGAGAGTACCTGCTGAGATATACCTGAACCTTGAGCGGCGCC 385
OY 281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 386 CAGCTTCTCCCTCTTATCTATGAGATTCAGGAAATTCATTCATTCAGATCA 445
OY 301 TyrIleMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThrGluLeuPro 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 446 TTTCACATGATGGACACAGATCATTCATGCTATTCATGAGATGACCACTTCC 505
OY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 506 CAGTGTGGCAATGATTAACCTTAAGAGTGCACCAATCATCAATTTAATTTACTTGAG 565
OY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 566 GAACATTTAAAAAACAAAGAGAAATTCGATCATATTCATACATGAGTACAGATGAGA 625
OY 361 AspiePheArgTyrArgHisSerValLysIleAsnGlyLysTyrAsnProGluValAsp 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 GGAAGAGAGAGTATACACACACTGTCATTAATGAGAGTGGACATCCAGAACTGAGC 685
OY 381 CysThrGlyLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 686 TGCTCAATGGCCAAATACATTAATGATGATGATGATGATGATGATGATGATGATGAT 745
OY 401 MetThrThrThrValAsnTyrGlnAspGlyLysValAlaValLeuCysLysGluAsn 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 746 ATGACAAACCACTGAATTAATCGGGATGAGAGAGAGATGCTCTTGGCCAAAGAAAT 805
OY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrGlnSerLeuPro 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 806 TATCTAATTCAGGAAGAGAAAGAAATTCATGCAAGATGAGAGAGTCAATACCA 865
OY 441 ArgCysValGlu----- 444
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 866 CTCTGTGTGAAGAAATTCATGTTCCACACCACTCAGATAGAACAGCAACATTAAT 925
OY 444 ----- 444
Db 926 TCATCCAGGCTTTCACAGAAAGTTATGACATGGAGCTAAATTGATTAATCTGTGAG 985
OY 444 ----- 444
Db 986 GGTGCTTCAGGATTCATGGAAGAAATGAAACACATGCTACATGGAAATGAGTTCT 1045
OY 444 ----- 444
Db 1046 CCACCTCAGTGAAGGCTTCTCTTAATCTCCACCGAGATTTCTCATGCTGTGTA 1105
OY 444 ----- 444
Db 1106 GCTCAGATGACAGACGTTATCAGTATGAGAGAGAGTACGTAACAAATGTTGAAGT 1165
OY 444 ----- 444
Db 1166 TTTGGAATTCATGGCCCTGCATTCGAAATGCTTAGAGAGAAATGCTTCACCCCTCCA 1225

```

```

OY 444 ----- 444
Db 1226 TCATGCATAAAAACAGATTCCTCAGTTACTTACTTACTTGAATAATGCCATGCCATGGGA 1285
OY 444 ----- 444
Db 1286 GAGAAAGAGATGTGTATTAAGCGGGGTGAGCAAGTACTTACCTTGTGCACAAATATTAC 1345
OY 444 ----- 444
Db 1346 AAAATGATGAGCCAGTAATGTAACATGATTAATAGCAGATGACAGAGAGCCCAACA 1405
OY 444 ----- 444
Db 1406 TGCAGACACCTCTCTGTGTAATCCGCCACAGTACAAAATGCTTATATAGTGCAGA 1465
OY 444 ----- 444
Db 1466 CAGATGATTAATATCCATCTGCTGAGAGAGTACGTTATCAATGATGAGAGCCCTTATGAA 1525
OY 444 ----- 444
Db 1526 ATGTTGGGGAATGAAGAGATGATGTTTAAATGGAACCTGACGCAACCACTCAATGC 1585
OY 445 ----- 445
Db 1586 AAAGATTTCTACAGAAATGCTGGGCCCTCCACCTATTCACAAATGGGACATTACTTCA 1645
OY 463 PheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSerPheTyr 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1646 TTCCTGTTGAGTATGATGCTCCACCTTATCATGAGTGGATGACCAATGCCAGAACTTGAT 1705
OY 483 LysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrSerGluProArg 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1706 CAACCTGAGGATACAAACCAATTAATGATGAGAAATGAGCAATGCTGCAACCAACAAA 1765
OY 503 CysLeuAspProCysValValSerGluGlnLysAsnMetLysAsnAsnIleGlnLeuLys 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1766 TGCTTACATTCGCTGTATATATCCGAGAAATTAATGAAATTAATACATGACATTAGG 1825
OY 523 TrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGlnCysLys 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1826 TGCAGACCCAAACAAACCTTTATTCGAGAACAGGTGATCAGTGAATTTCTGTGTAA 1885
OY 543 PheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGluLys 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1886 CGGGATATTCGCTTTCATCAGCTTTCACACATTGCCAACAACATGTTGGATGGAAA 1945
OY 563 PheGluTyrProIleCys 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1946 CTGAGATTCACCACTTGT 1963

RESULT 13
HUMHAB 1264 bp mRNA linear PRI 27-APR-1993
LOCUS Human factor H homologue mRNA, complete cds.
DEFINITION M65293
ACCESSION M65293.1 GI:183764
VERSION 1
KEYWORDS factor H.
SOURCE Human liver, cdna to mRNA, clone pFH.4b.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Estaller,C., Koistinen,V., Schwaebel,W., Dierich,M.P. and
Weiss,E.H.
Cloning of the 1.4-kb mRNA species of human complement factor H
reveals a novel member of the short consensus repeat family related
to the carboxy terminal of the classical 150-kDa molecule
J. Immunol. 146 (9), 3190-3196 (1991)
JOURNAL MEDLINE 91201892
PUBMED 1826708
FEATURES Location/Qualifiers

```


LOCUS BC016755 1296 bp mRNA linear PRI 09-NOV-2001
DEFINITION Homo sapiens, H factor (complement)-like 1, clone MGC:13525
IMAGE:3934474, mRNA, complete cds.
ACCESSION BC016755
VERSION BC016755.1 GI:16876960
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapps-remail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbioology.org>
contact: amadan@systemsbioology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAL Plate: 19 Row: a Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 11321586.
FEATURES
Source location/Qualifiers
1..1296
/organism="Homo sapiens"
/db_xref="LocusID:3078"
/db_xref="GI:16876961"
/clone="MGC:13525 IMAGE:3934474"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH-MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
93..1085
/codon_start=1
/product="H factor (complement)-like 1"
/protein_id="AAH16755.1"
/db_xref="GI:16876961"
/translation="MMLVSVILIRISVSGEATFCDFPRINHGILYDEEKYKPFQ
VPTGEVYVYCEVNFVSPKSFMTIRICTEGMSPTKCLCFPPVNGHSSGQ
THLEGDVOICNTGYLONNENISGVERGMSPPKCRSDTSCVNPATONAHILS
ROMSKYPSGKRVRECRSPYEMGDEDFVCMGMPKPPROCKSDSCGPPRPDND
ITSFPLSYAPASSVEYQCONLYOLEGNKRTCKNGMSEPKLHPCVISREIMENT
NIALRMTRAKOLVLTGESAFCVKRGVRLSSRSHTLNTTCMDKLEYPTCAKR"

Alignment Scores:
Pred. No.: 3.22e-85 Length: 1296
Score: 1054.50 Matches: 216
Percent Similarity: 43.49% Conservative: 31
Best Local Similarity: 38.03% Mismatches: 80
Query Match: 33.13% Indels: 241
DB: 9 Gaps: 5
US-09-805-337a-2 (1-569) x BC016755 (1-1296)
QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGlu 20
DB 93 ATGTGGCTGCTGTCAGTGAATCTATATCTACGAGATATCTCTGTTGGGGAGAACCA 152
QY 21 ThrLeuCyAspPheProLysIleHisHisGlyPheLeuTrpAspGluLysPTrasn 40
153 ACATTTTGATGATTTCCAAAATAAACCATGAGATTCATATGATGAAGAAATATTAAG 212
QY 41 ProPheSerGlnValProThrGlyGlnValPheLeuTrpSerCysGluTrpAsnPheVal 60
DB 213 CCATTTTCCACAGGTTCTACAGGGGAGATTTCATATACCTCTGTAATTAATTTGTG 272
QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGlnGluGluTrpSerProThr 80
DB 273 TCTCTTCCAAATATATTTTGGACGTCGATACATGACACAGAGAGATGTCACCAACA 332
QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlnHisSerGluSerSer 100
DB 333 CCAAGAGTGTCCACAGCTGTGTTCTCTTCCTTTGTGAAAATGTCATTCGATCTTCA 392
QY 101 G1yLeuIleHisLeuGlnGluGlyAspTrpValGlnIleLeuCysAsnThrGlyTrpSerLeu 120
DB 393 GGACAAACACATCTGGAGAGGATVACGTGCMAATTTTGCACACACAGCATACAGACTT 452
QY 121 GlnAsnAsnGlnLysAsnIleSerCysValGlnArgGlyTrpSerThrProProIleCys 140
DB 453 CAAACATGATGAGACACACATTTTCATGTGTAGAACGGGCTGTCCACCCCTCCAAATGC 512
QY 141 SerPheThrTrpLysGlyGluCysHisValProIleLeuGlnLysAsnValAspAlaGlnPro 160
DB 513 AGTCCACCTGAC----- 524
QY 161 LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
DB 524 ----- 524
QY 181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr 200
DB 524 ----- 524
QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGlnValLys 220
DB 525 -----ACTTCCTGTGATGATCCGCCACAGTACAAATGTCATATGACTG 569
QY 221 GlnLeuArgLysGlnGluTrpCylHisAsnGlnValValGluTrpAspCysAsnProAsn 240
DB 570 TCGAGACAGATGACTAATATCCATCTGGTAGAGAGACTGTATGAA----- 617
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro 260
DB 617 ----- 617
QY 261 ThrCysValGlnGlnValLysThrCysGlyTrpIleProGlnLeuGluTrpGlyVal 280
DB 617 ----- 617
QY 281 GlnProSerValProProTrpGlnHisGlyValSerValGlnValAsnCysArgAsnGlu 300
DB 618 -----TGTAAGAGCCCT 629
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGlnLeuPro 320
DB 630 TATGAATGTTTGGGAGTGAAGAGTATGATGCTTTAATGGAAGTGCACGAA----- 683
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
DB 683 ----- 683
QY 341 LeuLeuLysLeuSerGlyLysGlnPheAsnHisAsnSerArgIleArgTrpArgCysSer 360
DB 683 ----- 683
QY 361 AspIlePheArgTrpArgHisSerValCysIleAsnGlyLysTrpAsnProGlnValAsp 380
DB 683 ----- 683
QY 381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400

Db 684 -----CCACCTCAA----- 692
 QY 401 MetThrThrValAsnTyrGlnAspGlyGluValAlaValLeucysLysGluAsn 420
 Db 692 ----- 692
 QY 421 TyrLeuLeuProGluAlaLysGluLeuValCysLysAspGlyArgTyrGlnSerLeuPro 440
 Db 693 -----TGCAGAGAT----- 701
 QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460
 Db 702 -----TCTACGGGAAATGTGGGCCCTCCACCTATTGACAAATGGGACATT 749
 QY 461 ThrSerPheProLeuSerValTyrProGlySerThrValThrTyrArgCysGlnSer 480
 Db 750 ACTTCATTCCTCCGTTGTCAGTATATGCTCAGACTTCATCAGTGTAGTACCATGCCAGAAC 809
 QY 481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrPheSerGluPro 500
 Db 810 TTGATATCACTTGAGGGGTACACAGCAATGATGTAATGACATGCTCAGAACCA 869
 QY 501 ProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln 520
 Db 870 CCAAAATGCTTACATCCGTGTGTATATCCGAGAAATTATGAAATTTATACATAGCA 929
 QY 521 LeuLysTyrPArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540
 Db 930 TTAAAGTGGACAGCCCAACAGCAAGCTTATTTTGAGAGAGTGAATCAGCTGAATTGTG 989
 QY 541 CysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560
 Db 990 TGTAAACGGGATATCGCTTTTCATTCACGTTTCACACATTCGACACACATGTTGGAT 1049
 QY 561 GlyLysPheGluTyrProIleCys 568
 Db 1050 GGGAACTGGAGTATCCACTTGT 1073
 RESULT 15
 AX409739 1266 bp DNA linear PAT 14-JUN-2002
 LOCUS AX409739
 DEFINITION Sequence 2386 from Patent WO0229103.
 ACCESSION AX409739
 VERSION AX409739.1 GI:21442444
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1
 AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
 TITLE Gene expression profiles in liver cancer
 JOURNAL Patent: WO 0229103-A 2386 11-Apr-2002;
 GENE LOGIC INC (US)
 FEATURES
 source 1. 1266
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="EMBL/Genbank Accession No. M65292"
 BASE COUNT 414 a 237 c 249 g 366 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.13e-85 Length: 1266
 Score: 1050.50 Matches: 216
 Percent Similarity: 43.31% Conservative: 30
 Best Local Similarity: 38.03% Mismatches: 81
 Query Match: 33.00% Indels: 241
 DB: 6 gaps: 5
 US-09-805-337a-2 (1-569) x AX409739 (1-1266)
 QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTyrPheValSerThrValGlyGlyGly 20

Db 78 ATGCGCTCCCTGGTCACTGTAATCTCAATCTCAGCATATCTCTGTGGGGAGAGCA 137
 QY 21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTyrAspGluLysPheTyrAsn 40
 Db 138 ACATTTGTATTTTCCAAAATAAACCATGGAATTTATATATATGAAGAAATAATATAG 197
 QY 41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheVal 60
 Db 198 CCATTTTCCAGTCTCTACAGGGAGTTTTCATATACCTCTGTGAATTAATTTTGTG 257
 QY 61 SerProSerLysSerPheTyrPheThrArgIleThrCysThrGluGluTyrPheProThr 80
 Db 258 TCTCCCTCAAAATCATTTTGGACTCGCATCAATGACACAGAGAGATGTCACACACA 317
 QY 81 ProLysCysLeuTyrMetCysSerPheProPheValLysAsnGlyHisSerGluSer 100
 Db 318 CCAAGTGTCTCAGACTGTGTTCTTCTCTTTGTGAAATGTCAATTTGTAATCTTCA 377
 QY 101 GlyLeuIleHisLeuGluGluAspThrValGlnIleIleCysAsnThrGlyTyrSerLeu 120
 Db 378 GGACAAACACATCTGGAAAGTGATCTGCAATTTATTTGCAACACAGATACAGACT 437
 QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTyrPheThrProProIleCys 140
 Db 438 CAAACAAATGAGAACCAATTTTCATGTGTGAACGGGGCTGCCACCCCTCCCAATGC 497
 QY 141 SerPheThrLysGlyGlyCysHisValProIleGluGluAlaAsnValAspAlaGlnPro 160
 Db 498 AGGTCCACTGAC----- 509
 QY 161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
 Db 509 ----- 509
 QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTyrPheSerProAsnPheProThr 200
 Db 509 ----- 509
 QY 201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLys 220
 Db 510 -----ACTTCCTGTGTGAATCCGCCCACTACAAATATGTTATATAGTG 554
 QY 221 GluIleArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240
 Db 555 TCGAGACAGATGAGTAATATTCATCTGCTGAGAGATACGTTAT----- 559
 QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrThrLeuPro 260
 Db 559 ----- 559
 QY 261 ThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrGlyTyrVal 280
 Db 559 ----- 559
 QY 281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
 Db 600 -----CAATGCTAGAGCCCT 614
 QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyTyrPheThrGluLeuPro 320
 Db 615 TATGAATGTGTTGGGAGTGAAGAGTGAATGTTTAATGAAGAACTGAGCGGAA----- 668
 QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
 Db 668 ----- 668
 QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
 Db 668 ----- 668
 QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrPheAsnProGluValAsp 380

```
Db 668 ----- 668
QY 381 CysThrGluLysArgGluInpHecysProProProGlnIleProAsnAlaGlnAsn 400
Db 669 ----- 677
      |||||
      CCACCTCA-----
QY 401 MetThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
Db 677 ----- 677
QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrrpGlnSerLeuPro 440
Db 678 ----- 686
      |||||
      TGCAAGAT-----
QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460
Db 687 ----- 734
      |||||
      -TCTACGGGAAATGTGGGCCCCCTCCACCTATTGACAAATGGGACATT 734
QY 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
Db 735 ACTTCATTCCCGTTGTCAGTATATGCTCCAGCTTCATCAGTGTGAGTACCAATGCCAGAAC 794
QY 481 PheTyrLysLeuGlnIlySerValThrValThrCysArgAsnLysGlnTrrpSerGluPro 500
Db 795 TTGTATCAACTTGAGGGTAAAGCAGCAATTAACATGTAGAAATGACAAATGTCAGAACCA 854
QY 501 ProArgCysLeuAspProcysValValSerGluGluAsnMetAsnLysAsnAsnIleGln 520
Db 855 CCAAAATGCTTACATCCGTGTGTAATATCCCGAAATTTTGAAATTTAATACATRACA 914
QY 521 LeuLysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540
Db 915 TTAAAGGTGGACCAACCAAGAGCTTTATTGAGAAACAGTGAAATCAGCTCAATTGTCG 974
QY 541 CysLysPheProHisLysAlaMetIleSerSerProPheArgAlaIleCysGlnGlu 560
Db 975 TGTAAACGGGATATGTCGTCTTTCATCATCAGCTTCTCACACATTGCGAACACATGTGGGAT 1034
QY 561 GlyLysPheGluTyrProIleCys 568
Db 1035 GGGAAACTGGAGTATCCAACTGT 1058
```

Search completed: July 23, 2003, 10:49:57
Job time : 4214 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2003, 03:04:22 ; Search time 347 Seconds
(Without alignments)
3692.758 Million cell updates/sec

Title: US-09-805-337A-2
Perfect score: 3183
Sequence: 1 MLLFVILITSWSTVGEG.....SSPPFALCEGKFEYPICE 569

Scoring table:

BLOSUM62			
Xgapop 10.0 ,	Xgapext 0.5		
Ygapop 10.0 ,	Ygapext 0.5		
Dgapop 6.0 ,	Dgapext 7.0		
Delop 6.0 ,	Delext 7.0		

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n.model -DEV=xlh
-O/cqg2_1/USPTO.spool/US09805337/runat_17072003_135605_18885/app_query.fasta.1.711
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFX=ring -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdt
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09805337 @CGN.1.1.263 @runat_17072003_135605_18885 -NCPU=6 -ICPU=3
-NO.MMAP -LARGEIOUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -JONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3183	100.0	2823	AAD37114	Human factor H-rel
2	2733	85.9	1707	AAD37115	Human factor H-rel
3	2162	67.9	1225	ABK94924	Human novel polynu
4	1691	53.1	984	ABL57707	Human sbg614126com
5	1599	50.2	1095	ABL57706	Human sbg614126com
6	1303	40.9	2548	ABO55008	Human ovarian antl
7	1051.5	33.0	1293	AACT7947	Human cancer assoc
8	1050.5	33.0	1266	ABN58888	Gene #2386 used to
9	1001.5	31.5	2033	ABN54773	Human complement f
10	927.5	29.1	1040	ABN97258	Gene #3766 used to
11	756.5	23.8	2177	AAH82402	B subunit of human
12	710	22.3	1269	ABN97281	Gene #3779 used to
13	681.5	21.4	1929	AA506029	Angiotensin conver
14	681	21.4	1315	ABN97342	Gene #3840 used to
15	515	16.2	767	AAV02792	Human partial comp
16	487	15.3	759	AAK39793	Gastric cancer ass
17	466	14.6	688	AAV02794	Clone pBS3FH2910
18	432.5	13.6	590	AAV02795	B lymphocyte membr
19	424	13.3	3327	AAQ10989	Human secreted pro
20	395	12.4	2929	AA503887	Receptor 222 codin
21	395	12.4	3696	AAH75787	Human gene GS96663
22	395	12.4	3706	ABN93421	Human cDNA SEQ ID
23	395	12.4	3804	ABAO6548	Human cDNA encoding for
24	395	12.4	3804	AA528968	Human cDNA encoding
25	395	12.4	3804	AA526868	Human cDNA encoding
26	395	12.4	3804	AA531587	Human cDNA encoding
27	395	12.4	3804	ABK43685	Human REPR 12 cDN
28	395	12.4	3991	ABK13101	Human C3b/C4b comp
29	394	12.4	10878	AAD32025	Angiotensin conver
30	389.5	12.2	1194	AA506052	Human POLY11 nucle
31	389	12.2	6153	ABAO3880	Human C3b/C4b rece
32	385	12.1	5420	AAQ11643	Human C3b/C4b rece
33	385	12.1	5420	AAZ38151	Human C3b/C4b rece
34	385	12.1	5420	AB165862	Lung cancer relate
35	385	12.1	5420	ABN1637	Human C3b/C4b rece
36	385	12.1	6951	AAH91477	Entire human compl
37	385	12.1	6951	AAQ11642	Human C3b/C4b rece
38	385	12.1	6951	AAZ38150	Human cDNA differe
39	385	12.1	6951	ABK84738	Human C3b/C4b rece
40	385	12.1	6951	ABK81636	Human polynucleoti
41	385	12.1	7313	AA158380	DNA encoding novel
42	385	12.1	7821	AA564474	DNA encoding novel
43	385	12.1	9038	AA564290	Mouse C3b/C4b comp
44	384.5	12.1	11230	AAD32026	Human MDT encodin
45	383.5	12.0	2575	ABQ72635	

ALIGNMENTS

RESULT 1
ID AAD37114 standard; cDNA: 2823 BP.
AC AAD37114;
DT 21-AUG-2002 (first entry)
DE Human factor H-related protein (FHR-5) encoding cDNA.
KW Human: factor H-related protein; FHR-5; detection; C5b-9 complement;
KW Immunohistochemical; diagnostic; immunological; biopsy; prevention;
KW gene; ss.
XX Homo sapiens.
OS
XX
FH Key
FT CDS Location/Qualifiers
94..1803

FT /tag- a
FT /product- "Human factor H-related protein (FHR-5)"
PN MO200168695-A2.
XX
XX 20-SEP-2001.
PD 13-MAR-2001; 2001WO-0507868.
PF 13-MAR-2000; 2000US-0188870.
PR (BAXT) BAXTER INT INC.
XX
XX Murphy BF;
XX MPI: 2001-582437/65.
DR P-PSDB: AAE23195.
XX
XX
PT New factor-H related protein 5 that binds to complement component C3b,
PT useful for raising antibodies suitable for detection of C3b-9
PT complement complexes,
PS Claim 3; Page 50-52; 61pp; English.
XX
XX The invention relates to a novel human factor H-related protein
CC (FHR-5), its corresponding nucleic acid and antibodies directed
CC to it. The antibody is useful for detecting C3b-9 complement
CC complexes, by contacting a cell or tissue with Ab and detecting
CC the binding of Ab to the cell or tissue. Ab is useful as
CC immunohistochemical diagnostic reagent to detect immunological
CC deposits in biopsied tissues, and also for preventing the
CC association of FHR-5 with activated complement. The present
CC sequence is human FHR-5 cDNA.
XX
SQ Sequence 2823 BP; 960 A; 531 C; 546 G; 786 T; 0 other;
Alignment Scores:
Pred. No.: 1,666-272 Length: 2823
Score: 3183.00 Matches: 569
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-805-337A-2 (1-569) x ADJ37114 (1-2823)
QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
DB 94 ATGTGCTCTTATTGAGTGAATCCAACTCATGCGTATCCACTGTGGGAGAGAGA 153
QY 21 ThrLeuCysAspPheProLysIleHisIsglyPheLeuTrpAspGluGlyTrpAsn 40
DB 154 ACCTTTTGATTTTCCAAAATAACACCATGGATTTCTATGATCATAGAAAGATTATAC 213
QY 41 ProPheSerGlnValProThrGlyGluValPheTrpTyrSerCysGluTrpAsnProVal 60
DB 214 CCTTTTCCCAAGTTCTACAGGGAGATTTCTATTACTCCGTGAATATATATTTG 273
QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr 80
DB 274 TGTCCCTCAAAATCCCTTTGAGCTGCATACATGCAGAGAGAAAGATGTCACCAACA 333
QY 81 ProLysCysLeuArgNgcCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100
DB 334 CCGAAGTGTCTCAGATGTGTCCTTCTCTTGTGAAAAATGGTCATCTGAATCTTCA 393
QY 101 GlyLeuIleHisLeuGluGluLysPheValGlnIleIleCysAsnThrGlyTrpSerLeu 120
DB 394 GGACTAATACATCTCGAAGGTGATACTGTAACAAATATTGCAACACAGATACAGCCCT 453
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys 140
DB 454 CAAACAAATGAGAAAAACATTTTCGTGTGTAGAAAGGGGCTGGTCCACTCTCCCATATGC 513

QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluValAsnValAspAlaGlnPro 160
DB 514 AGCTTCACCTAAGAGAAATGTCAATGTTCCAAATTTTGAAGCCCATGTACATGCTGACCA 573
QY 161 LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
DB 574 AAAAAAGAAAGCTACAAAGTTGACAGCTGTGCAAAATTTCTCCGACGAAAAATCTTATA 633
QY 181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnProThr 200
DB 634 AGAGTTGATCAGACTCAGCTTCATGTTACCAATTTGGGTGGTCACTTAACCTTCCAAACA 693
QY 201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyValLys 220
DB 694 TCGAAGAGCAAGTAGCATCATGTGTCCACCTCCCTCAACTCTCCCAATGTTAG 753
QY 221 GlnIleArgLysGluGluTrpGlyHisAsnGluValGluTrpAspCysAsnProAsn 240
DB 754 GAGATAGAGAAAGAGAAATATGACCAATGAGTGAATATGATGCAATCTTAAT 813
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260
DB 814 TTTATATATAACGGCTTAAGAAATACATGTGTGATGCAATGACAACTTACC 873
QY 261 PheCysValGluGlnValLysThrCysGlyTrpIleProGluLeuGluTrpGlyTrpVal 280
DB 874 ACTTGCTGCAACAACTGAAACATGTGATACATCTGAACTGACATGACGATGTT 933
QY 281 GlnProSerValProProTrpGlnHisIsglyValSerValGluValAsnGlyArgGlu 300
DB 934 CAGCCCTGCTCCCTCCCTCATCAACATGAGTTTCAGTGCAGTGAATTCGCAAAATGAA 993
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrLeuPro 320
DB 994 TATGCAATATGATGGAATATACATGATTACTGTATTAAAGATATGACAGACGTTCT 1053
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
DB 1054 ATGTGTTGCAACACACCACCACTTAAGAGTGCAAAATAGCAGGATTAATTA 1113
QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTrpArgCysSer 360
DB 1114 TTACTCAACCTATCTGGGAGAGAAATTAATCAATTAATCTAGAAATGCTTACAGATGTTCA 1173
QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
DB 1174 GACATCTCAGATACAGGCACTCAGTCTGTATTAACGGGAATGGAATCTCAAGTAGAC 1233
QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400
DB 1234 TGCACAGAAAAAGGCAAAATTCCTCCACCGCCACCTCAATACCTAATCTCAGAAAT 1293
QY 401 MetThrThrThrValAsnTrpGlnAspGlyGluLysValAlaValLeuLysLysGluAsn 420
DB 1294 ATGACACACACAGTGAATATATACAGATGAGAAAAAGTACGCTTCTCTGTAAGCAAAAC 1353
QY 421 TyrLeuLeuProGluValLysGluIleValCysLysAspGlyArgTrpAsnLeuPro 440
DB 1354 TATCTACTTCCAGGAAGCAAAAGAAATTTATATGAAGATGAGCAATGGCAATCATTACCA 1413
QY 441 ArgCysValGluSerThrAlaTrpCysGlyProProProSerIleAsnAsnGlyAspThr 460
DB 1414 CGCTGTGTGAGTCACTCATATGTTGTGGCCCTCCATCTTAATTAACATGAGATATAC 1473
QY 461 ThrSerPheProLeuSerValTrpProProGlySerThrValThrTrpArgGlyGlnSer 480
DB 1474 ACCGTATTCCTCAT 1533
QY 481 PheTrpLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500
DB 1534 TTCTATAAATCTCAGGCGCTGTACTAATCAATGACAGAAATTAACAGTGGTCAAGACCA 1593

QY	501	ProaGcysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln	520
Db	1594	CCAAAGATCCCAAGATCCATGCTGTGTAATCTGAAGAAACATGCAACAAAATATACATACAG	1653
QY	521	LeuLysTSPATGAsnAsnPGlyLysLeuLysAlaLysThrGlyAspAlaValGluPheGln	540
Db	1654	TTAAATATGAGAAACGATGCGAAACTATATGCAAAACAGGGGATGCTGTGATTCGAG	1711
QY	541	CysLysPheProHisLysLysAlaMetLysSerSerProPheArgAlaIleCysGlnGlu	560
Db	1714	TGTAAATTCGCCACATAAAGCGATGATATCATCCACACCATTTTCGAGCAATCTGTCAGGA	1771
QY	561	GlyLysPheGluTyrTyrProIleCysGlu	569
Db	1774	GGGAATTTGAAATATCCTATATGTGAA	1800
RESULT 2			
ID	AAD37115	standard; DNA; 1707 BP.	
XX	AAD37115;		
AC	21-AUG-2002	(first entry)	
XX			
DE	Human factor H-related protein (FHR-5) degenerate DNA.		
XX			
KW	Human; factor H-related protein; FHR-5; detection; C5b-9 complement;		
KM	immunohistochemical; diagnostic; immunological; biopsy; prevention;		
XX	gene; ds.		
OS	Homo sapiens.		
XX			
PN	WO200168695-A2.		
XX			
PD	20-SEP-2001.		
XX			
PE	13-MAR-2001; 2001WO-US07868.		
XX			
PR	13-MAR-2000; 2000US-0188870.		
PA	(BAXT) BAXTER INT INC.		
XX			
PI	Murphy BF;		
XX			
DR	WPI; 2001-582437/65.		
PT	New factor-H related protein 5 that binds to complement component C3b,		
PT	useful for raising antibodies suitable for detection of C5b-9		
PT	complement complexes,		
PS	Claim 3; Page 54-55; 61pp; English.		
XX			
CC	The invention relates to a novel human factor H-related protein		
CC	(FHR-5), its corresponding nucleic acid and antibodies directed		
CC	to it. The antibody is useful for detecting C5b-9 complement		
CC	complexes, by contacting a cell or tissue with Ab and detecting		
CC	the binding of Ab to the cell or tissue. Ab is useful as		
CC	immunohistochemical diagnostic reagent to detect immunological		
CC	deposits in biopsied tissues, and also for preventing the		
CC	association of FHR-5 with activated complement. The present		
CC	sequence is human FHR-5 degenerate DNA.		
XX			
SO	Sequence 1707 BP; 340 A; 170 C; 280 G; 231 T; 686 other;		
Alignment Scores:			
Pred. No.:	7,51e-233	Length:	1707
Score:	2733.00	Matches:	475
Percent Similarity:	83.48%	Conservative:	0
Best Local Similarity:	83.48%	Mismatches:	94
Query Match:	85.86%	Indels:	0
DB:	23	Gaps:	0

QY	I	MelleuleuleupheserValIleleulesertPvalSerThrValGlygluglucy	20
Db	1	ATGTNTNTNTNTTWSMCGTNATHYTNTAHTSNMGGSNMCNGTNGGNCAGCAGCN	60
QY	21	ThrlLeuCysApplpheProLysIleHisIsglyPheLeuTyraSpGlugluAsnTyraSn	40
Db	61	ACNTNTNGCYGATTYCCNAARATHCATCAVGCTTTTNTATYGATGARGARATTAAAY	120
QY	41	PropheSerGlnValProThrglyGluValPheHtyrTySerCysGluTyraSnPheVal	60
Db	121	CCHTYSWNCARGTTCNCACNGCAGCARTNTTNTATYAWSMTGYGARTAAVATYTGTN	180
QY	61	SerProserLySerPheTrpThrArgIleHrcysThrGluGluGlyTTPSerProthr	80
Db	181	WSNCCMSMAARWSMTTYTGACNMGNATHCAGTNCNGARGRAGGNTGGMSNCCACN	240
QY	81	ProltyScysLeuArgMetCysSerPheProPheAllysasnGlyHisSerGlnSerSer	100
Db	241	CCMARITGYTNMGATHTGTGWSMTTYGCNTTGTGNNARAAYGGNCAYSGMSAGAWSMNS	300
QY	101	GlyLeuIleHisLeuGluGlyAspThrValGlnIleileCysasnThrGlyTyrrSerLeu	120
Db	301	GGNYTNATHCAAYTTGARGGNGAAVCANCTNCARATHTGTGAAYACNGGTAWSNVYN	360
QY	121	GlnAsnaSnGluTyraSnIleSerCysValGluArgrgGlyTTPSertPrroProlieCys	140
Db	361	CARAAAYAGARAARAAYATHTWSNTGTGTNCARRKNGGNTGGMSNCCGCCCATHTGX	420
QY	141	SerPheThrLySglGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGluPro	160
Db	421	WSMTTYACNAARCGCARTGYCAVGTGCCNCAHTYTNGARGCANAAYGTNGAAGCCNARC	480
QY	161	LysLySglUsetTyrrLySValGlyAspAlleuleysPheSerCysArgLysAsnLeuIle	180
Db	481	AARARAGRMSNTATAAARGTNGCAGCATNTNTNAARTTWSNTGYMGAAAAYTNATH	540
QY	181	ArgValglySerAspSerValGlcysTryGlnPheGlyTTPSerProasnPheProThr	200
Db	541	MONGTNGMSNGAWSMNGTNCARTGYATYACARTTYGGNTGMSNCCAATVTYYCCNANC	600
QY	201	CysLySglGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLys	220
Db	601	TGYAARGNCARGLTMGMWSMTGYGNCNCNCNCNCARYTNWSNAAAGGAGAGCTNAAR	660
QY	221	GluIleArgLysGlnGlyTyrglyHisasnGluValValGluTyraSPcysAsnProsn	240
Db	661	CARATHMGMAARGAGARTATGGNCATYAATYCARGTNGTNGAKRTAYGATGTAAATCCNAAY	720
QY	241	PheIleIleasnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeupro	260
Db	721	TTYATHAHTHAAYGGCNCNAARAARATHCARTGYGTNGAYGGNGARTGACNACVYTNCCN	780
QY	261	ThrcysValGluGlnValLysThrcysGlyTyrrIleProGluLeuGluTyrglyTyraVal	280
Db	781	ACNTGYIGNAGCACGTINAARCACTGYGNTATYATHCNGARGRYTNGATTAAGSTPATRGTN	840
QY	281	GlnProSerValProProTyrglnHisGlyValSerValGluValAsnCysArgAsnGlu	300
Db	841	CARCCMSNGTNCNCNTAVCARCAVGGNTNMNSNGTNGARGTNAAYTGYMGNAAYGAR	900
QY	301	TyrAlaMetIleGlyAsnAsnMetIleHrcysIleasnGlyIleTrpThrGluLeupro	320
Db	901	TAYGCMATGATHGGAAYAAATGATHCAGTATHAAYGGNAHTGTGACAGAYTTCNCN	960
QY	321	MetCysValAlaThrHisGlnLeuAsnArgCysLysIleAlaGlyValAsnIleLysThr	340
Db	961	ATGTGTGTCNCACACATCARTYTNAARKNGTGYAARATHGCGNGGCTAAATATPAARACN	1020
QY	341	LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleatgyTyraCySer	360
Db	1021	YTNYNTHARLYTWMSGCNAARGARTTYAACYAVALYWSNMOMNATHTMGTAATGMYGTWNSN	1080

OY	361	AspIlePheArGlyrArGHisSerValCysIleAsnGlyLysTrpAspProGluValAsp	380
Db	1081	GAATGATHTTTCMGNTATAMGCATWMSNTNGVATHTAAVGGNAATGGAAATCCGARTNSAY	1140
OY	381	CysThrGluLysArGcGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn	400
Db	1141	TGYACNCGARRARMGAGCARTTGTGTGCGCCNCCNCCNARATHCCNAAYGCCNCAARAY	1200
OY	401	MetThrTrpThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn	420
Db	1201	ATGACNACNACNGTNAATTAATCAARATGAGNARAAAGTNGTNTGTGAAARARAY	1260
OY	421	TyrLeuLeuPProGluAlaLysGluIleValCysLysAspGlyArGTrpGlnSerLeuPro	440
Db	1261	TAYTNTNTNCCNGARCGCAARAGARTHTGTNTGYARAGVAGCGMGNTGGCARSNTNCCN	1320
OY	441	ArGcysValAlGluSerThrAlaTyrCysGlyProProPProSerIleAsnAsnGlyAspThr	460
Db	1321	MGNTGCGTNGNARWMSNACNGCNTATGTGCGNCCNCCNMSNATHTAAAYAAVGGNGATVACN	1380
OY	461	ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArGcysGlnSer	480
Db	1381	ACNMSNTTCCNTYNTWMSNGTNTATVCCNCCNCGNMSNACNGTNACTATATMGNTGYCARSN	1440
OY	481	PheTyrLysLeuGlnGlySerValThrValThrCysArGAsnLysGlnTrpSerGluPro	500
Db	1441	TTYTAYAAARYNCAARGGMSNGTNAACNCTNACNTGMYMNAAYAAARCATGCGMSNARCCN	1500
OY	501	ProArGcysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln	520
Db	1501	CCNMGNTGTYNTNGATCCNTGTGCTNCTWMSNCGARCARAAATATGAATAAAATAAATVTHCAR	1560
OY	521	LeuLysTrpArGAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValAlGluPheGln	540
Db	1561	YTNAARTCGMNAAYAGAGYAGNAARTNTATYGCNNAARACNGNCGATCGNTGCARTTYCAR	1620
OY	541	CysLysPheProHisLysAlaMetIleSerSerProProPheArGAlaIleCysGlnGlu	560
Db	1621	TGYAARTTCCNCAVYAARGCNATGATWMSNMSNCCNCTTGMCGCATHTGTGCARGAR	1680
OY	561	GlyLysPheGluTyrProIleCysGlu 569	
Db	1681	GGNARTTGYARATVCCNATHTGTGAR 1707	
RESULT 3			
ID	ABK94924	standard; cDNA: 1225 BP.	
AC	ABK94924:		
XX	30-AUG-2002	(first entry)	
DT	30-AUG-2002		
XX			
DE	Human novel polynucleotide #35.		
XX			
KW	Human; gene; ss; inflammatory condition; shock; sepsis; immune response;		
KW	cancer; wound healing; central nervous system disease; haematopoiesis;		
KW	peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;		
KW	myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;		
KW	cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;		
KW	bone degenerative disorder; periodontal disease; reperfusion injury;		
KW	lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;		
KW	allergic condition; thrombolysis; thrombosis; coagulation disorder;		
KW	fungal infection. *		
XX			
OS	Homo sapiens.		
XX			
PN	W020024340-A2.		
XX			
PD	06-JUN-2002.		
XX			
PF	30-NOV-2001; 2001WO-US47004.		
XX			
PR	30-NOV-2000; 2000US-0028952.		

XX	(HISE-) HXSEQ INC.
PA	Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI	Yamazaki V, Ujwal ML, Drmanac RT;
XX	WPI: 2002-508509/54.
DR	P-PsDB: ABG66700.
XX	Novel nucleic acids and polypeptides for diagnosis, treatment of
PT	inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT	disorders, cancer and promoting wound healing
PS	Claim 1; Page 406-407; 672pp; English.
XX	The invention relates to human novel polynucleotides and associated
CC	polypeptides. The polynucleotides and polypeptides are useful for
CC	treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC	disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC	and cancer and for promoting wound healing. The sequences are used to
CC	induce the proliferation of neural cells and regeneration of nerve and
CC	brain tissue, and are useful for the treatment of central and peripheral
CC	nervous system diseases and neuropathies, such as Alzheimer's disease,
CC	Parkinson's disease, Huntington's disease and amyotrophic lateral
CC	sclerosis. The sequences are involved in chemotactic or chemokinetic
CC	activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC	cell disorders and platelet disorders such as thrombocytopenia,
CC	regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC	growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC	osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC	disease. The sequences of the invention are also useful for gut
CC	protection or regeneration and treatment of lung or liver fibrosis,
CC	reperfusion injury in various tissues, immune deficiencies and disorders
CC	including severe combined immunodeficiency (SCID), bacterial or fungal
CC	infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC	gravis, allergic conditions such as asthma, thrombolytic or thrombolysis
CC	and coagulation disorders. Sequences ABK94890-ABK94982 represent human
CC	novel polynucleotides of the invention.
SQ	Sequence 1225 BP; 425 A; 243 C; 243 G; 314 T; 0 other;
XX	
XX	
Alignment Scores:	
Pred. No.:	2,4e-182 Length: 1,225
Score:	2162.00 Matches: 384
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	67.92% Indels: 0
DB:	Gaps: 0
US-09-805-337A-2 (1-569) x ABK94924 (1-1225)	
OY	186 SerValGInCysTyrGlnIlePheGlyTrpSerProAsnPhcProThrCysLysGIgInVal 205
Db	24 TCAGTTCAATGATTACCAATTTGGGTGGTGTCACCTACTTCCACACATGCAGAACAGTA 83
OY	206 ArgSerCysGlyIleProProGluLeuSerAsnGIgInValLysGIgInIleArgLysGIu 225
Db	84 CGATCACGTGTCACCCTCCACACTTCCTCAATGGTGAAGTAAGGAGATAAGAAGAGAG 143
OY	226 GIuTrfGIyHisAsnGIuValValGIuTrfAspCysAsnProAsnPhelIleAsnGIy 245
Db	144 GAATATGACACAATGAACTAGTAGGAGATATGATTCATCTCAATTTTATATATTAACGGG 203
OY	246 ProLysLysIleGInCysValAspGIgInTrpThrIleauProThrCysValGIgIn 265
Db	204 CCTAAGAAAATACAAATGCCGTGATGAGCATGCGAACACTTATGCCACCTTGTTGAACAA 263
OY	266 ValLysThrCysGIyTrfIleProGluLeuGIuTrfGIyTrfValGIuInProSerValPro 285
Db	264 GTGAATACATGTGGATACATACCTGACTCGAGTACGGTATATGTTCACGCCGTGTCTCC 323
OY	286 ProTyrGIuInIsGIyValSerValGIuValAsnCysArgAsnGIuTrfAlaMetIleGIy 305

```

Db      324 CCTATCAACATCGACTTCAGGTGGAATTCAGAAATGATATGCAATGATTGCA 383
Oy      306 AsnAsnMetIleThrCysIleasnGlyIleTrpHgluLeuProMetCysValAlaThr 325
Db      304 AATACATGATTACTGTTATTTATGGAATATGCAAGACTTCCTTATGTGTGTTGCACA 443
Oy      326 HisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeuLeuLysLeuSer 345
Db      444 CACCAACTTAAGAGGTGCAGAAATAGCAGAGCTTAATATAAAACAATTACTCAAGCATCT 503
Oy      346 GlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAspIlePheArgTyr 365
Db      504 GGGAAAGAAATTTAACTATTAATCTAGAAATACCTTACAGATGTCAGACATCTTCAGATAC 563
Oy      366 ArgHisSerValCysIleasnGlyLysTrpAsnProGluValAspCysThrGlyLysArg 385
Db      564 AGGCATCTAGTGTGTTATTAACGGGAAATGGATCCCTGAAGTGAAGTGCACAGAAATAAG 623
Oy      386 GluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMetThrThrThrVal 405
Db      624 GAACAATCTGCCACACCGCCACCTCAGATACCTAATGCTCAGAAATATGACAAACACAGT 683
Oy      406 AsnTyrGlnAspGlyGlyLysValAlaValLeuCysLysGluAsnTyrLeuLeuProGlu 425
Db      684 AATTATCAGGATGAGGAAAAAGTACCTGCTCTCTTAAAGAAACCTATCTCTCCAGAA 743
Oy      426 AlaLysGluIleValCysLysAspGlyArgTyrGlnSerLeuProArgCysValGlnSer 445
Db      744 GCAAAAGAAATTTGTTATGTAAGATGAGCATGGCAATCTTACCCAGCTGTGTTGAGTCT 803
Oy      446 ThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThrThrSerPheProLeu 465
Db      804 ACTGCATATTTGGGCCCCCTCCATCTATTACAAATGAGATACCACTCATCTCCATTA 863
Oy      466 SerValTyrProProGlySerThrValThrTyrArgCysGlnSerPheTyrLysLeuGln 485
Db      864 TCAGTATATCCCTCCAGGGTCAACAGTACGCTGTCGACGCTCTCTTATAACTCCAG 923
Oy      486 GlySerValThrValThrCysArgAsnLysGlnTyrSerGluProProArgCysLeuAsp 505
Db      924 GGCTCTGTATCTGTAACATGCAAGAAATAAACAGTGTGTGAAACCCACCAAGATGCCAT 983
Oy      506 ProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGlnLeuLysTrpArgAsn 525
Db      984 CCATGTGTCGATCTGAGAGAAACAATGAACAATAATACATCAGTTAAATGAGAGAAAC 1043
Oy      526 AspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGlnCysLysPheProHis 545
Db      1044 GATGAAAAAAGCTTATGCAAAACACGGGATGCTGTTGAATTCAGTAAATCCCAAT 1103
Oy      546 LysAlaMetIleSerSerProProPheArgAlaIleCysGlnGluGlyLysPheGluTyr 565
Db      1104 AAAGCATATATCTATCATCACACCATTTTCAGCAATCTGTACAGAGGAAATTTGATAT 1163
Oy      566 ProIleCysGlu 569
Db      1164 CCTATATGTGAA 1175

```

RESULT 4

```

ABL57707
ID      ABL57707 standard; DNA; 984 BP.
AC      ABL57707;
XX      08-OCT-2002 (first entry)
XX      Human sbg614126comp1fh gene #2.
XX      Human; secreted protein; immunosuppressive; cytostatic; neurotropic;
KW      neuroprotective; antitumor; vulnerrary; antimicrobial; ophthalmological;
KW      antiparkinsonian; antileumatic; antiatherosclerotic; dermatological;
KW      hypotensive; cerebroprotective; vitucide; antiinflammatory; diabetes;
KW      malignant tumor; hypertension; hypotension; obesity; bulimia; anorexia;

```

```

KW      asthma; manic depression; dementia; delirium; mental retardation;
KW      Huntington's disease; Tourette's syndrome; schizophrenia;
KW      mental disorder; sexual development disorder; blood cascade dysfunction;
KW      stroke; growth disorder; gene; ds.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      CDS      1..984
FT      /tag=a
FT      /product="sbg614126comp1fh"
XX      WO200222802-A1.
XX      21-MAR-2002.
XX      13-SEP-2001; 2001WO-US28462.
XX      13-SEP-2000; 2000US-232455P.
XX      13-SEP-2000; 2000US-232463P.
XX      02-OCT-2000; 2000US-237283P.
XX      07-NOV-2000; 2000US-246269P.
XX      20-NOV-2000; 2000US-252049P.
XX      (SMK ) SMITHKLINE BECHAM CORP.
XX      (SMK ) SMITHKLINE BECHAM PLC.
XX      (GLAX ) GLAXO GROUP LTD.
XX      Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smuth RF, Strum JC;
XX      Xie Q;
XX      WPI; 2002-393963/42.
XX      P-PSDB; ABB80571.
XX      Novel isolated secreted polypeptides and polynucleotides encoding them
XX      useful for treating cancer, Alzheimer's disease, tumor metastasis,
XX      autosomal recessive atypical hemolytic uremic syndrome, wound healing
XX      disorder
XX      2: Page166; 246pp; English.
XX      CC The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)
XX      CC (secreted polypeptide) which is encoded by any one of 44 polynucleotide
XX      CC sequences (ABL57705-ABL57748) given in the specification. The
XX      CC polypeptides have immunosuppressive, cytostatic, neurotropic,
XX      CC neuroprotective, antitumor, vulnerrary, antimicrobial, ophthalmological,
XX      CC antiparkinsonian, antileumatic, antiatherosclerotic, dermatological,
XX      CC hypotensive, cerebroprotective, vitucide, and antiinflammatory activity.
XX      CC The polynucleotide and polypeptide are useful for treating diabetes,
XX      CC malignant tumours, hyper- and hypotension, obesity, bulimia, anorexia,
XX      CC asthma, manic depression, dementia, delirium, mental retardation,
XX      CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
XX      CC or sexual development disorders, and dysfunctions of the blood cascade
XX      CC system including those leading to stroke.
XX      SQ Sequence 984 BP; 309 A; 201 C; 194 G; 280 T; 0 other;

```

Alignment Scores:

```

pred. No.: 1,18e-140      Length: 984
Score: 1691.00      Matches: 326
Percent Similarity: 57.47%      Conservative: 1
Best Local Similarity: 57.29%      Mismatches: 0
Query Match: 53.13%      Indels: 242
DB: 24      Gaps: 2

```

US-09-805-337A-2 (1-569) x ABL57707 (1-984)

```

Oy      1 MetLeuLeuLysPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
Db      1 ARGTTGCTTTATTCAGTGTATCCAAATCTCATGGTATCCACGTCCTTGGGAGAGAGA 60
Oy      21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTyrAspGluAspTyrAsn 40

```

```

Db      61 ACACCTTGATTTCCAAAATACACCATGATTTCTGTATGATGAGAGATTATTAAC 120
Qy      41 ProPheSerGlnValProThrGlyGluVal1PheTYrTYrSerCysGluTYrAsnPheVal 60
Db      121 CCTTTTCCCAATCTCTACAGGGAGATTTCATTAATCTACTCCGTGATAATATATTTTGG 180
Qy      61 SerProSerLysSerPheThrPheArgIlePheCysTYrGlnGluGlyTYrSerProThr 80
Db      181 TCCCTTCAAAATCCCTTTGACTCCGATACATGCACAGAGAGAGATGTCACCAACA 240
Qy      81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100
Db      241 CCGAATGCTCTCAATGCTGTTCTTCTTCTGTAATAAAGTGATATTCGATCTTCA 300
Qy      101 GlyLeuIleHisLeuGlnGlyAspThrValGlnIleIleCysAsnThrGlyTYrSerLeu 120
Db      301 GGACTATATACATCTCGAAGGTGATCTAGTACAATAATATTGCAACAGAGATACACCTT 360
Qy      121 GlnAsnAsnGlnLysAsnIleSerCysValGluArgGlyTYrPheThrProPheIleCys 140
Db      361 CAATAACATAGAGAAAACATTTCTGTGTAGAACGGGGCTGCTCCACTCCCATATGC 420
Qy      141 SerPheThrLysGlyGlyCysHisValProIleLeuGlnAlaAsnValAspAlaGlnPro 160
Db      421 AGCTTCACT----- 429
Qy      161 LysLysGluSerTYrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
Db      429 ----- 429
Qy      181 ArgValGlySerAspSerValGlnCysTYrGlnPheGlyTYrPheProAsnPheProThr 200
Db      429 ----- 429
Qy      201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLys 220
Db      429 ----- 429
Qy      221 GluIleArgLysGluGluTYrGlyHisAsnGluValValGluTYrAspCysAsnProAsn 240
Db      429 ----- 429
Qy      241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTYrPheThrLeuPro 260
Db      429 ----- 429
Qy      261 ThrCysValGluGlnValLysThrCysGlyTYrIleProGlnLeuGluTYrGlyTYrVal 280
Db      430 -----ATGAATAACATGTGATACATACCTGCACTCGAAGTACGAGTTATGTT 474
Qy      281 GlnProSerValProProTYrGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
Db      475 CAGCCCTGCTCCCTCCCTATCAACATGAGATTTCTGAGAGTGAATGCGAATAATGAA 534
Qy      301 TYrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTYrPheGlnLeuPro 320
Db      535 TATGCATATATTTGGAATAATACATGATTAACCTGTATTATGAGATATGACAGAGCTTCT 594
Qy      321 MetCysValAlaIleThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
Db      595 ATG----- 597
Qy      341 LeuLeuLysLeuSerGlyLysGlnPheAsnHisAsnSerArgIleArgTYrArgCysSer 360
Db      597 ----- 597
Qy      361 AspIlePheArgTYrArgHisSerValCysIleAsnGlyLysTYrAsnProGlnValAsp 380
Db      597 ----- 597
Qy      381 CysThrGlnLysArgGlnGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
Db      597 ----- 597

```

```

Qy      401 MetThrThrThrValAsnTYrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
Db      597 ----- 597
Qy      421 TYrLeuLeuProGlnAlaLysGluIleValCysLysAspGlyArgTYrGlnSerLeuPro 440
Db      597 ----- 597
Qy      441 ArgCysValGluSerThrAlaTYrCysGlyProProProSerIleAsnAsnGlyAspThr 460
Db      598 ---TGTTGAGTCTTACATATGATGAGGCCCCCTCCATCTATTATTAACATGAGATACC 654
Qy      461 ThrSerPheProLeuSerValTYrProProGlySerThrValThrTYrArgCysGlnSer 480
Db      655 ACCTCATTCCTCCATTATACATATATTCCTCCAGGGTCAACAGTGACGTACCGTCCAGTCC 714
Qy      481 PheTYrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTYrPheGluPro 500
Db      715 TTCTATTAACCTCCAGGGCTGTGTACTGTAAACATGCGAGAAATAACAGTGTCCAGAACCA 774
Qy      501 ProArgCysLeuAspProCysValValSerGlnGluAsnMetAsnLysAsnIleGln 520
Db      775 CCAAGATGCCCTTGATTCATGTGTGTATCTGAAGAAACATACACAAATATACATACAG 834
Qy      521 LeuLysTYrPheArgAsnAspGlyLysLeuTYrAlaIleValThrGlyAspAlaValGlnPheGln 540
Db      835 TTAATAATGCAGAAACAGATGAAACCTATGCAAAAACAGGGAGTGTGTAATTCAG 894
Qy      541 CysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560
Db      895 TGTAAATTCACACATTAACGATGATATCATACACACATTTCGAGCAATVCTGTAGGAA 954
Qy      561 GlyLysPheGluTYrProIleCysGlu 569
Db      955 GCGAAATTTGAAATATCCTATATGTGAA 981

RESULT 5
ABLS7706
ID      ABL57706 standard; DNA; 1095 BP.
XX
AC      ABL57706;
XX
DT      08-OCT-2002 (first entry)
XX
DE      Human sbg614126complfH gene #1.
XX
KW      Human; secreted protein; immunosuppressive; cytostatic; neutrotropic;
KW      neuroprotective; antitumour; vulnerrary; antimicrobial; ophthalmological;
KW      antiparkinsonian; antirheumatic; antiatherosclerotic; dermatological;
KW      hypotensive; cerebroprotective; virucide; antiinflammatory; diabetes;
KW      malignant tumour; hypertension; hypotension; obesity; bulimia; anorexia;
KW      asthma; manic depression; dementia; delirium; mental retardation;
KW      Huntington's disease; Tourette's syndrome; schizophrenia;
KW      mental disorder; sexual development disorder; blood cascade dysfunction;
KW      stroke; growth disorder; gene; ds.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..1095
FT      FT      /tag=a
FT      FT      /product="sbg614126complfH"
XX
PD      WO200222802-A1.
XX
XX      21-MAR-2002.
XX
XX      13-SEP-2001: 2001WO-US28462.
XX      PF
XX      13-SEP-2000: 2000US-232455P.
XX      PR
XX      13-SEP-2000: 2000US-232463P.
XX      PR
XX      02-OCT-2000: 2000US-237293P.
XX      PR

```

PR 07-NOV-2000: 2000US-246269P.
 PR 20-NOV-2000: 2000US-252049P.
 XX
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 PA (SMIR) SMITHKLINE BEECHAM PLC.
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smith RF, Strum JC,
 PI Xie Q;
 XX
 DR WPL: 2002-393963/42.
 DR P-PDB: ABB80570.
 XX
 PT Novel isolated secreted polypeptide and polynucleotides encoding them
 PT useful for treating cancer, Alzheimer's disease, tumor metastasis,
 PT autosomal recessive atypical hemolytic uremic syndrome, wound healing
 PT disorder
 XX
 PS 2; Page166; 246pp; English.
 XX
 CC The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)
 CC (secreted polypeptide) which is encoded by any one of 44 polynucleotide
 CC sequences (AB157705-AB157748) given in the specification. The
 CC polypeptides have immunosuppressive, cyostatic, nootropic,
 CC neuroprotective, antitumor, vulnerary, antimicrobial, ophthalmological,
 CC antiparkinsonian, antirheumatic, antithrombotic, dermatological,
 CC hypotensive, cerebroprotective, virucide, and antiinflammatory activity.
 CC The polynucleotide and polypeptide are useful for treating diabetes,
 CC malignant tumors, hyper- and hypotension, obesity, bulimia, anorexia,
 CC asthma, manic depression, dementia, delirium, mental retardation,
 CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
 CC or sexual development disorders, and dysfunctions of the blood cascade
 CC system including those leading to stroke.
 XX
 SQ Sequence 1095 BP: 341 A: 217 C: 223 G: 314 T: 0 other:

 Alignment Scores:
 Pred. No.: 2.03e-132 Length: 1095
 Score: 1599.00 Matches: 307
 Percent Similarity: 56.00% Conservative: 1
 Best Local Similarity: 55.82% Mismatches: 0
 Query Match: 50.24% Indels: 242
 DB: 24 Gaps: 2

 US-09-805-337A-2 (1-569) x AB157706 (1-1095)
 QY 20 GlyThrLeuGlySerPheProLysIleHisGlyPheLeuTyraSpGluAspTyr 39
 DB 169 GGAACCTTTGGATTTTCAGAAATACACCATGGATTCTGTATGAGAGATTTAT 228
 QY 40 AsnProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPro 59
 DB 229 AACCTTTTCCCAAGTCTCTACAGGGGAAGTTTCTATTACTCCGTGAATTAATTTT 288
 QY 60 ValSerProSerLysSerPheTyrPThrArgIleThrCysThrGluGluTyrProPro 79
 DB 289 GGTGTCCTTCAAAATCCTTTTGACCTCCATTAACATGACAGAAAGAGATGTCACCA 348
 QY 80 ThrProLysCysLeuArgMetCysSerPheProPheValLysAsnGlnHisSerGluSer 99
 DB 349 ACACCCAAATGCTCTGGAATGTGTCTCTTCTTTGTGAAAATGGTCAATCTGATCT 408
 QY 100 SerGlyLeuIleHisLeuGluGlyAspThrValGlnIleLeuCysAsnThrGlyTyrSer 119
 DB 409 TCAGGACTATATACATCTCGAAGGTGATGATGTAACAATTAATTCGAACACAGATACAG 468
 QY 120 LeuGlnAsnAsnGlnLysAsnIleSerCysValGluArgGlyTyrProProIle 139
 DB 469 CTTCAAAACAAATGAGAAAAACATTTGTGTGTAGAAAGGGGCTGCTCCACTCTCCATA 528
 QY 140 CysSerPheThrLysGlyGluCysHisValProIleLeuGlnLysAsnValAspAlaGln 159
 DB 529 TGCAGCTTCACT----- 540

QY 160 ProLysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeu 179
 DB 540 ----- 540
 QY 180 IleArgValGlySerAspSerValGlnCysTyrGlnPheGlyTyrProSerProAsnPro 199
 DB 540 ----- 540
 QY 200 ThrCysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGlnVal 219
 DB 540 ----- 540
 QY 220 LysGluIleArgLysGluGluTyrGlyHisAsnGlnValValGluTyrAspCysAsnPro 239
 DB 540 ----- 540
 QY 240 AsnPheIleTLeAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrLeu 259
 DB 540 ----- 540
 QY 260 ProThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrGlyTyr 279
 DB 541 -----ATGAACATGTGATACATACCTGAACCTCGAGTACGCTTAT 582
 QY 280 ValGlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsn 299
 DB 583 GTTACACCGCTGTCTCCCTCCATACATGAGATTCAGTTCAGTGCAGTATTCACCAAT 642
 QY 300 GluTyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThrGluLeu 319
 DB 643 GATATGCAATGATTGGAAATTAACATGATTACTGTATTAAAGGATATGACAGAGCTT 702
 QY 320 PrometCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLys 339
 DB 703 CCTATG----- 708
 QY 340 ThrLeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCys 359
 DB 708 ----- 708
 QY 360 SerAspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrTrpAsnProGluVal 379
 DB 708 ----- 708
 QY 380 AspCysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGln 399
 DB 708 ----- 708
 QY 400 AsnMetThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGlu 419
 DB 708 ----- 708
 QY 420 AsnTyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrProLysLeu 439
 DB 708 ----- 708
 QY 440 ProArgCysValGlnSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAsp 459
 DB 709 -----TGTTGTAGTCTACTGATATTTGTGGCCCCCTCCATCTTAATTAACATGGAGAT 762
 QY 460 ThrThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGln 479
 DB 763 ACACCTCATTTCCCATTAATATATATCTCCAGGGTCAACAGTACGTAACCGTTGCCAG 822
 QY 480 SerPheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrProGlu 499
 DB 823 TCCCTCTATTAACCTCCAGGGCTCTGTAACTGTAAACATGAGAAATTAACAGTGGTCAGAA 882
 QY 500 ProProArgCysLeuAspProCysValIleSerGluGlnAsnMetAsnLysAsnIle 519
 DB 883 CCACCAAGATGCTATGATCATGATGTGTGTGAGAGAAACATGAAACAAATTAACATA 942

Db 948 AAGAAGATGATACACACACTCTGCATTAATGGAAGATGGATCCAGACAGTGAATGCTC 1007
 QY 382 ThrGluysArgGluInphecysPropProProGlnIleProaspNlaGlnAsnMet 401
 Db 1008 TCATATGGCCAAATATACATTAATGCTCCACCTCCACCTCCACATTCCTCCATTCATATG 1067
 QY 402 ThrThrValAsnTyrGlnAspGlyGluysValAlaValLeuGlyGlnAsnTyr 421
 Db 1068 ACAACACACTGAATATATGCGATGAGAGAAAAGTATCTGTTCTTGGCCAGAAAATATAT 1127
 QY 422 LeuLeuProGluAlaIleValGlyGlnAspGlyArgTyrGlnSerLeuProArg 441
 Db 1128 CTATATTCAGGAAGAGAGAAATATACATCAAAAGATGGAAGATGCGATCAATACCACTC 1187
 QY 442 CysValGlu----- 444
 Db 1188 TGTGTGAAAAAATTCATGTTCCACACACCTCAGATAGAACCGGAACCATTAATTCGA 1247
 QY 444 ----- 444
 Db 1248 TCCAGGCTTTCACAGAAGATTAATGCATGAGCACTAATTAATGATTACTTGTGAGGCT 1307
 QY 444 ----- 444
 Db 1308 GGTTCAGATATCTGGAAGAAAAATGAACAACATGCTACATGGAATAATGAGTTCTCCA 1367
 QY 444 ----- 444
 Db 1368 CCTCAGTGTGAAGGCTCTCTTGTAAATCTCCACCTGACATTTCTCATGGTGTGTAGCT 1427
 QY 444 ----- 444
 Db 1428 CACATGTCAAGAGTATCATGATGAGAGAAAGTACGTAACAAATGTTTGAAGSTTTT 1487
 QY 444 ----- 444
 Db 1488 GGAATTGATGGGCTGCAATTCGCAAAATGCTTAGAGAAAAAATGTCACCTTCATGCA 1547
 QY 444 ----- 444
 Db 1548 TGCATAAAAAGATTTGTCTCAGTTTACCTAGCTTGAATAATGCCATACCATGGAGAG 1607
 QY 444 ----- 444
 Db 1608 AAGAAGATGTATATAAGCGGGGTGAGCAAGTGCATTTGCAACATATTACAAA 1667
 QY 444 ----- 444
 Db 1668 ATGGATGAGCCAGTATATGATACATCATTAATAGCAGATGAGAGGAGCCACATGCG 1727
 QY 444 ----- 444
 Db 1728 AGAGACACTCTCTGTGTGATCCGCCACAGTACAAAATGCTTATATAGTGTGAGACAG 1787
 QY 444 ----- 444
 Db 1788 ATGAGTAATATTCATCTGTGAGAGATAGCTTATCATATGAGAGCCCTTATGAATG 1847
 QY 444 ----- 444
 Db 1848 TTGGGGATGAAGAGTATGTGTTTAAATGGAAGCTGAGACGACACCTCATGSCAAA 1907
 QY 445 ---SerThrAlaTyrCysGlyProProSerIleAsnAsnGlyAspThrSerPhe 463
 Db 1908 GATTCACAGAAAATGTGGGCCCCCTCCACCTATTGACATGGGACATTAATTCATTC 1967
 QY 464 ProLeuSerValTyrProGlySerThrValThrTyrArgCysGlnSerPheTyrLys 483
 Db 1968 CCTTGTGAGTATATCTCCAGCTCATGATGATGACCAATGACAGAACTGTATCA 2027
 QY 484 LeuGlnGlySerValThrValThrCysArgAsnGlnTyrSerLeuProProArgCys 503
 Db 2028 CTTGAGGGTAAACAAGCAATATACATGTAGAAATGAGCAATGGTCAAGACCAACCAAAATGC 2087

QY 504 LeuAspProCysValIleSerGluIleAsnMetAsnLysAsnIleGlnLeuLysTrp 523
 Db 2088 TTATCATCCCTGTGTATATCCCGAGAAATATATGAGAAAATTAATACATTAAGTGG 2147
 QY 524 ArgAsnAspGlyLysLeuTyrAlaIleValThrGlyAspAlaValGluInphecInCysLysPhe 543
 Db 2148 ACAGCCAAACAGACCTTATTCGAGAACAGGTGATGAGTGAATTTGTGTAAACGG 2207
 QY 544 ProIleLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGluLysPhe 563
 Db 2208 GGAATATGCTTTTCATCAGCTTCTCCACATTCGGAACAAACATGTTGGATGGAGAAACGTG 2267
 QY 564 GluTyrProIleCys 568
 Db 2268 GAGTATCCAACTTGT 2282
 RESULT 7
 AAC77947 standard; cDNA: 1293 BP.
 AAC77947:
 AC
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated gene sequence SEQ ID NO:341.
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
 KW antidiabetic; antistatic; antihemetic; antitumor; antiviral;
 KW antineoplastic; antihypertensive; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 KW immunologic; antiproliferative; antineoplastic; gene therapy; inflammation;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.
 OS Homo sapiens.
 XX
 PN MO20005350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587533/55.
 XX
 DR P-PSDB: AAB43738.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 XX
 PS Claim 1; Page 892; 2352pp; English.
 XX
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnary; immunomodulator;
 CC antidiabetic; antistatic; antihemetic; antitumor; antiviral;
 CC antineoplastic; antihypertensive; antiallergic; antibacterial; cardiant;
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC neotropic; vasotropic; antiproliferative; antineoplastic; gene therapy;
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 OS Homo sapiens.
 PN WO200229103-A2.
 XX
 PD 11-APR-2002.
 PF 02-OCT-2001: 2001WO-US30589.
 XX
 PR 02-OCT-2000: 2000US-237054P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX WPI: 2002-426119/45.
 XX
 PT Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample
 PS
 PS Claim 1: SEQ ID NO 2386; 298pp; English.
 XX
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcc_sequences.
 XX
 SQ Sequence 1266 BP; 414 A; 237 C; 249 G; 366 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 1,27e-83 Length: 1266
 Score: 1050.50 Matches: 216
 Percent Similarity: 43.318 Conservative: 30
 Best Local Similarity: 38.034 Mismatches: 81
 Query Match: 33.004 Indels: 241
 DB: 24 Gaps: 5
 US-09-805-337A-2 (1-569) x ABN95888 (1-1266)
 QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
 DB 78 ATGTGCTCTCGTGTAGTGAATTCATCTCAGCATATCCCTGTTGGGGAGAGCA 137
 QY 21 ThrLeuGlyAspPheProLysIleHisGlyPheLeuTyraAspGluLysAspTyrasn 40
 DB 138 ACATTTTGTGATTTTCCAAAATAAACCATGCAATTCATATGATGACAAAAATATAG 197
 QY 41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheVal 60
 DB 198 CCAATTTTCCAGGTTCTCTACAGGGAGATTTCCTATTAACCTCGTCAATATATTTTGTG 257
 QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluLysTyrPseProThr 80
 DB 258 TCTCCCTTCAAAATCATTTTGGACTGCGATACATGCACAGAGAGAGATGTCACCAACA 317
 QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100

DB 318 CCAAGTGTCTCAGACTGTGTTCTTCTTTGCAAAAAGTCATTCGAATCTTCA 377
 QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAspThrGlyTyrSerLeu 120
 DB 378 GGACAAACACATCTGGAGGTGATCTGTCAAAATTTATTTGACACACAGATACACTT 437
 QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys 140
 DB 438 CAAACCAATGACACACACATTTTCATGTGTGAGAACGGGTGGTCCACCCCTCCAAATGC 497
 QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160
 DB 498 AGCTCCACTGAC----- 509
 QY 161 LysLysGluSerTyrLysValGlyLysPheValLeuLysPheSerCysArgLysAsnLeuIle 180
 DB 509 ----- 509
 QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnPheProThr 200
 DB 509 ----- 509
 QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLys 220
 DB 510 -----ACTTCTGTGTGATCCGCCACAGTACAAATGCTTATATAGTG 554
 QY 221 GluIleArgLysGluGluLysGlyHisAsnGluValGluTyrAspCysAsnProAsn 240
 DB 555 TCGAGACATGATGATTAATATTCATCTGCTGAGAGAGATGTTAT----- 599
 QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260
 DB 599 ----- 599
 QY 261 ThrCysValGlnGlnValLysThrCysGlyTyrIleProGluLeuGluLysGlyTyrVal 280
 DB 599 ----- 599
 QY 281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
 DB 600 -----CAATGTAGGAGCCCT 614
 QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyTyrTrpThrGluLeuPro 320
 DB 615 TATGAATGTGTTGGGATGACAGAGATGCTTTAAATGGAACGACGGA----- 668
 QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
 DB 668 ----- 668
 QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
 DB 668 ----- 668
 QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
 DB 668 ----- 668
 QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400
 DB 669 -----CCACCTCA----- 677
 QY 401 MetThrThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
 DB 677 ----- 677
 QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro 440
 DB 678 -----TGCAAAAGAT----- 686
 QY 441 ArgCysValGlnSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460


```

Db      780 AGAGTCGAGTACGAGGCTCTACTGAACTTCAGGCTCTCAAAATGTAACATGT 839
QY      252 VALASPGLYGLUTRPHRTHLEUPROTHRYVALGLUGLVALLYSTHRCYSGLYTYR 271
Db      840 ACTAATGGAGACTGGCAACCAAGATGCATA--TCAATGAACCTTGAGTTT 896
QY      272 ILEPROGLULEUGLUTRGY-----TYRVALGLPROSERVALPROFOTYR----- 287
Db      897 ---CCAGAAATTCACATGACATCATATTTATGAAATACCGCTAGACATATCTTCCA 953
QY      288 ---GLNHISGLYVALSERVALGLUVALAANCYSARGASGLUTYRALAMELLEGLYASN 306
Db      954 GTAGCTACAGCAACTCTACTCTATTTACTGTGACCAAAATTTTGTGACTCTTCAGGA 1013
QY      307 -----AsmMetLeuHrcysIleasnGLYIleTyrPhrGLUleuProMetCysVal 323
Db      1014 ACTTACTGGATTTACTTCTACTGTCACACAGATGGGTGGTCCAAAGCTCCCATC--- 1070
QY      324 ALATHRHISGLILEULYSARGCYSYLIleAGLYVALASNILEYSTHR--LeuLeu 342
Db      1071 -----CTCAGAACATGCTTCACAAATGACATATGAAATGAAATGATTCATT 1118
QY      343 LYSLEUSERGLYLYSGLUHPHASNHISASNSEARGLIleRGTYRARGYSSERASPILIE 362
Db      1119 TCTGAATCTCTCTATTTATTTAAATTAATAAATATCAATATTAATTAACACAGGA 1178
QY      363 PHEARGTYRARGHISER-----ValCysIleasnGLYIleYSTPASPAN 376
Db      1179 TATGCACACGACAGATGAAATCTTCAGGTTCAATTACATGTTGGCAAAATGATGCTCA 1238
QY      377 PROGLUVALASPCYSTHRGLULYSARGGLUGLHPHecysProProProProGLINILEPRO 396
Db      1239 GCACACCAATTTGCATT-----AAATTTGT---CATATGCTGTTTTCAG 1283
QY      397 ASNALAGLASNMECTRTHRTHRVAlASNTRYGLIASPGLYGLULYSVALAlAVILLEU 416
Db      1284 AATTCACAGCCAGAGATAGGATGCGCGTTTAACTCCATGACACATTCGACTGCAGAA 1343
QY      417 CYSLYSGLIASNTRYLEUPROGLUVALALYS-----GLULYleALYSLSYASAP 433
Db      1344 TGTCTACGATGATATATAAATCAGTTTGGAAACACCAAGGTTCCATAGTGTGTGTGGA 1403
QY      434 GLYARGTRPGLNSELEUPROARGYVALGLUSERTHRAlATYRCYSGLYPROPPOPRO 453
Db      1404 GATGGGTGGCCCATTTCCACACATGTTPAANTTTTCAGAAAGAGTGGCCCTCTCCA 1463
QY      454 SERILEASNANGLYSPHTRTHRSERPHROLEUSERYALTYRPROPOGLYSERTHR 473
Db      1464 CCTATTAGCAATGGTATACACACCTCTTCTACTAAAGTGTATGTGCCACAGTCACAGA 1523
QY      474 VALTHTYRARGCYSGLINSERPHEUHYLYLYLEUGLINSERVALTHRVAlTHRYASRG 493
Db      1524 GTTCGAGTACCAATGCGCAGTCTTACTGTAACTTCAGGGTTCTAATATGATGACATGTA 1583
QY      494 ASNLVSGLITRPSERGLUPROPROARGYCYLEUASPPROCYVALVALSERGLUGLUSN 513
Db      1584 AATGGAGAGGTGGCGGAGACCAACAGATGCATATCATATGATATTAATCTGAAGAAAC 1643
QY      514 METASNLVYASNANLLEGLINLEULYSTPRARGASNSPGLYLYSLEUTYRALALYSTHR 533
Db      1644 ATGAATAAATAAATACATACAGTTAAAGAAAAGACATAAATAATATATGCAAAAACA 1703
QY      534 GLYASVALAlVALGLUPHegLINCYSLYSPHeproHISLYSAlAMELLESERSEPRPRO 553
Db      1704 GGGGATACCATTAATTTATGTATTAATTGGGATATAAATGCGAATCATCATGCTTATCA 1763
QY      554 PHEARGAlAlIECYSGLINGLUGLYSPHegLUTYRPROILECYSGLU 569
Db      1764 TTTCAGCAGTGTGTAGGAGGCATAGTACCCAGATCGCA 1811

```

RESULT 10
ABN97268

```

ID      ABN97268 standard; DNA; 1040 BP.
XX
AC      ABN97268;
XX
DT      13-AUG-2002 (first entry)
XX
DE      Gene #3766 used to diagnose liver cancer.
XX
KW      Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW      metastatic liver tumor; cytostatic; expression profile; disease state;
KW      disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS      Homo sapiens.
XX
PN      WO200229103-A2.
XX
PD      11-APR-2002.
XX
PF      02-OCT-2001; 2001WO-US30589.
XX
PR      02-OCT-2000; 2000US-237054P.
XX
PA      (GENE-) GENE LOGIC INC.
XX
PI      Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX      WPT; 2002-426119/45.
XX
PT      Diagnosing and detecting the progression of liver cancer,
PT      hepatocellular carcinoma or metastatic liver tumor in a
PT      liver tissue sample -
XX
PS      Claim 1; SEQ ID NO 3766; 298bp; English.
XX
CC      The invention relates to a novel method for diagnosing and detecting the
CC      progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC      tumor in a patient, and differentiating metastatic liver cancer from
CC      hepatocellular carcinoma in a patient, involving detecting the level of
CC      expression of two or more genes represented in ABN93503-ABN97455 in a
CC      tissue sample. The method of the invention has hepatotropic, and
CC      cytostatic activity. The method is useful for diagnosing and detecting
CC      the progression of liver cancer, hepatocellular carcinoma and metastatic
CC      liver carcinoma in a patient. The method is useful for identifying
CC      expression profiles which serve as useful diagnostic markers as well as
CC      markers that can be used to monitor disease states, disease progression,
CC      drug toxicity, drug efficacy and drug metabolism.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SO      Sequence 1040 BP; 348 A; 192 C; 185 G; 315 T; 0 other;
XX
Alignment Scores:
Pred. No.:      8.07e-73      Length:      1040
Score:          927.50      Matches:     198
Percent Similarity: 38.11%      Conservative: 20
Best Local Similarity: 34.62%      Mismatches: 47
Query Match:    29.14%      Indels:      307
DB:             24      Gaps:        3
US-09-805-337A-2 (1-569) x ABN97268 (1-1040)
QY      1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
Db      78 AGTGGCTCTCTGCTCGATTAATTCACGAGATATCTTGGGGAAGA 137
QY      21 ThrLeuCysAspPheProLysIleHISGLYPhLeuTYRAspGLUAspTYRAsn 40
Db      138 ATGTTCTGTGATTTTCCAAAATAAACATGATTCATATGATATAAGAAAATATTAAG 197
QY      41 ProPheSerGLUValProThrGLYGLUValPheTYRYSerCysGLUTyrAsnPheVal 60

```

```

Db 198 CCAATTTTCCCAAGTTCCTACAGGGAAGTTTCTATTACTCTCGTAATATATTTGTG 257
QY 61 SerProSerLysSerPheThrPthrArgIleThrCysThrgluGluGlyTrpSerProthr 80
Db 258 TCTCCTTCAAAATCCTTTTGGACTCGCATACGTGGCAGAGAGAGATGTCACACACA 317
QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGlySer 100
Db 318 CCAAGGTCTCAGACTGTCTTCTTCTTCTTGTGGAAGAAATGTCATCTGATCTTCA 377
QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu 120
Db 378 GGACAAACACATCTGGAAGGTATCTATCAAAATATTTCACACAGATACAGACT 437
QY 121 GlnAsnAsnGlyLysAsnIleSerCysValGluArgGlyTrpSerThrProProLys 140
Db 438 CAAACAAATGACACACAACTTTCATGTATGAGAACGGGGTGGTCCACTCTCCCAATGC 497
QY 141 SerPheThrLysGlyLcysHisValProIleLeuGlnAlaAsnValAspAlaGlnPro 160
Db 497 ----- 497
QY 161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
Db 497 ----- 497
QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnProThr 200
Db 497 ----- 497
QY 201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyLysValLys 220
Db 497 ----- 497
QY 221 GluIleArgLysGluLcylTrGlyHisAsnGluValAlaGluTyrAspCysAsnProAsn 240
Db 497 ----- 497
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260
Db 497 ----- 497
QY 261 ThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGlnLcylTrGlyVal 280
Db 497 ----- 497
QY 281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnGlyAsnGlu 300
Db 497 ----- 497
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320
Db 497 ----- 497
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
Db 497 ----- 497
QY 341 LeuLeuLysLeuSerCylLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
Db 497 ----- 497
QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
Db 497 ----- 497
QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400
Db 497 ----- 497
QY 401 MetThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuLysCysLysGluAsn 420
Db 497 ----- 497

```

```

QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro 440
Db 497 ----- 497
QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460
Db 498 AGTCACACATTTCTGCAGAAAAATGTGGGCCCTCCACCATATGACAAATGAGACATT 557
QY 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
Db 558 ACTTATTCCTGTTGTCAAGTATATCTCTCCAGGTTATCATGAGTACCGACAGCCAGAAC 617
QY 481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500
Db 618 TTGTATCAACTTGCAGAGTACAAATCAATATACATGTAGAAACGCAATGTGCACAAACA 677
QY 501 ProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln 520
Db 678 CCAAAATGCTTAGATCCATGCTATATATCACAAGAAATTATGAAAAATATACATTAAA 737
QY 521 LeuLysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540
Db 738 TTAAGTGGACAAACCAACAAAGCTTTATTTCAAGACAGGTGACATGATTGATTGTT 797
QY 541 CysLys-----PheProHisLysAlaMetIleSerSerProProPheArgAlaIle 557
Db 798 TCTAATCTGATATCATCCACAA-----TTCATTCATTTCGACCAATG 845
QY 558 CysGlnGluGlyLysPheGluTyrProIleCysGlu 569
Db 846 TGTCAAAATGGGAACGTGTATATCCAGTTGTGAG 881

RESULT 11
AAN82402
ID AAN82402 standard; DNA; 2177 BP.
XX
AC AAN82402;
XX
DT 26-NOV-1990 (first entry)
XX
DE B subunit of human Factor XIII.
XX
KW B sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;
KW crosslink; scleroderma; haemorrhage; ulcerative colitis; ss.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT CDS 2..1521
FT FT /*tag= a
FT FT /product= b subunit
FT FT misc_RNA 2..58
FT FT /*tag= b
FT FT /label= leader_sequence
FT FT CDS 59..1521
FT FT /*tag= c
FT FT /product= mature b subunit
PN A08778694-A.
PD 31-MAR-1988.
XX
PF 21-SEP-1987; 87AU-0078694.
XX
PR 19-SEP-1986; 86US-0909512.
XX
PA (ZYMO-) ZYMOGENETICA INC.
XX
PI Davie E, Seale RL, Ichinose A, Holly JA, Parker GE;
XX WPI: 1988-140637/21.
XX
DR P-PSDB: A082921.

```

XX New DNA sequences encoding sub-units of factor 13 - and corresponding
PT expression vectors and transformed host cells.
XX

XX Disclosure: ; P; English.

XX The carboxyl-terminal Thr (nucleotides 1979-1981) is followed by
CC a stop codon (TAG), a 187 bp noncoding sequence, and a poly(A) tail
CC of 9 bp. The polyadenylation or processing signal of AATAAA
CC was identified 19 nucleotides upstream from the poly(A) tail.
CC When cultured the host cells will produce the polypeptide which
CC can be assembled to factor 13, which stabilises blood clots by
CC crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.
CC in cases of scleroderma, haemorrhage, ulcerative colitis etc., and
CC can be prep. in large aunts. without risk of viral contamination.
CC See also AAN82401 and AAN82403.

XX Sequence 2177 BP; 758 A; 358 C; 434 G; 627 T; 0 other;

Alignment Scores:

Pred. No.:	3,48e-57	Length:	2177
Score:	756.50	Matches:	194
Percent Similarity:	41.46%	Conservative:	85
Best Local Similarity:	28.83%	Mismatches:	235
Query Match:	23.77%	Indels:	139
DB:	9	Gaps:	21

US-09-805-337A-2 (1-569) x AAN82402 (1-2177)

QY 3 LeuLeuPheSerValIleLeuIleSerTTPValSerThValGlyGlyGlyThrLeu 22
DB 14 CTGACTTTATCATCATATTGATA-----ATCTCAGAGAA----- 49
QY 23 CysAspPheProLysIleHisGlyPheLeuTyrAspGluLysPyrAsnProPhe 42
DB 50 -----CTCTATGCAAGAGAGAAACCTGTGCTTT 79
QY 43 SerGlnValProThrGlyGlyVal-----PheTyrTyrSer----- 54
DB 80 CCTCATGTGGAATAATGGAAGAAATGGCCCATATTACTATCTTTAAAGCTTTACTTT 139
QY 55 -----CysGluTyrAsnPheValSerProSer 63
DB 140 CCAATGAGCATACACAAAAATTTGTCATTTTCTGCTGGCTGGTATATACCACTGAAGT 199
QY 64 LysSerPheThrPheThrGlyLeuThrCysThrGluGlyTyrSerProThrProLysCys 83
DB 200 GGAAGACAAAGAGAGCAACACGCTGTACACAGAAAGGCTGCTCTCCAGGCCAAGGTGC 259
QY 84 LeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSerGlyLeuIle 103
DB 260 TTCAAAAAATGACACTAAGCCTGACCTGAGTAATGTTACATCTCTGATGAAGTTATTG 319
QY 104 His----- 104
DB 320 TATTAATTCAGAGAACATGATATATGTTGGCTTCAGAGGTACAAACCACTGAGAGG 379
QY 105 -----LeuGluGly----- 107
DB 380 AAGGATGAAGAAAGTGGTTCAATCTCTCTGATGATGTTCTTCTCAACCAACCTGAGG 439
QY 107 ----- 107
DB 440 AAGAACATGAACGTTTGGCTCGATTAATTAATGAATTAATTCACACACAG 499
QY 108 -----AspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGln 121
DB 500 AAACATTCACAACTGAAGCAACAAAGTACATACGAAATGCTGCTACTGCTACACAGCT 559
QY 122 AsnAsnGluLys-----AsnIleSerCysValGluArgGlyThrSerThrProPhe 139
DB 560 GGAAGAAAGAAAGCAGAGAGGTAGAAATGTTCACTATGAGATGTTCTTCTCACACCAAAA 619

QY 140 CysSerPheThrLysGlyGlyCysHisValProIleLeuGluAlaAsnValAspAlaGln 159
DB 620 TGT-----ACCAATTAAGTGTCTCTTAAAGTATATGAAATGCTTATTTTCAT 673
QY 160 ProLysLysGluSerTyrLysValGlyAspValLeuLysPheSerArgLysAsnLeu 179
DB 674 CTGTAAACCAAAACCTATGACAGAGAGATGCTTCAGTTTCTTCTCATGAAATATAT 733
QY 180 IleArgValGlySerSerPheValGlnCysTyrGlnPheGlyTyrSerProAsnPro 199
DB 734 TATCTAAGTGATCTGATTAATTAATCATCTTAATCTTGGTGGTACCAAGATCTCT 793
QY 200 ThrCysLysGlyGlnValAlaArgSerCysGlyProProGluLeuSerAsnGlyVal 219
DB 794 GTATCGGAAGAAAGAAAGAAAGATGCTCTCCCTCCACCTCTGCCCAATAACCTCAAAAT 853
QY 220 LysGluIleArgLysGluGlyGlyHisAsnGluValGluTyrAspCysAsnPro 239
DB 854 ---CAACACATTCACAACTTATCTCATGAGAAATAGTCTATGAAATCTGAACCT 910
QY 240 AsnPheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluThrThrLeu 259
DB 911 AATTTTGCATTCATGCTGCTGACGACAGAAATACGTTGAAAGGTGGAATGAGACAGACT 970
QY 260 ProThrCysValGlu-----GlnValLysThrCysGlyTyrIleProGluLeuGluTyr 277
DB 971 CCAAAATGCATGTAGACAGACAGAAAGTACCTGTGAGGAAACCACTTCATGTAAT 1030
QY 278 GlyTyrValGlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCys 297
DB 1031 GGTGACAGCAAAATTTACACTTAAGATTTATTAACAAGGGGATTAACATGACATGATG 1090
QY 298 ArgAsnGluTyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThr 317
DB 1091 AAAGCGGTACTCTTCTCCATGATGATCAATGACATGATCTTGTATCTGGAATGACAA 1150
QY 318 GluLeuProMetCysValAlaThrHisGlnLeuLysArgCysLys----- 332
DB 1151 CTTCCTCTGACGAGTGTGTGAATAATATGAG-----AATGTATAGCATCTCTCTGTGTA 1204
QY 333 IleAlaGlyValAlaAsnIleLysThrLeuLysLeuSerGlyLysGluPheAsnHisAsn 352
DB 1205 ATGAATGGGCTGCTTCAGACGCGGATATTG-----GCAAGCTATGCAACAGCA 1252
QY 353 SerArgIleArgTyrArgCysSerAspIlePheArgTyrArgHisSerVal----- 369
DB 1253 TCCTCAGTGAATATGATGATGATGATATATCTTACTGAGGAGGATCAAAAATATCTCGT 1312
QY 370 CysIleAsnGlyLysThrPasnProGluValAspCysThrGluLysArgGluGlnPheCys 389
DB 1313 TCGGACAAAGAAATGTCATCCCACTGTTGCTTGGAAACA-----TGT 1360
QY 390 ProProProGluIleProAsnAlaGlnAsnMetThrThrValAsnTyrGln----- 408
DB 1361 ACTGTTAATGTGATTAACATGACAGAAATATACATGAATGGAATATATGAAGG 1420
QY 409 -----AspGlyGlyLysValAlaValLeuCysLysAsnTyr----- 421
DB 1421 AAAGCTTACATGAGATTAATATAGATTTTGTATGTAACAGGAGATATGACTTATCTCA 1480
QY 422 LeuLeuProGluAlaLys-----GluIleValCysLysAspGlyLysPheProGluLeuPro 440
DB 1481 TTAAACCAATTTGCTGATATATGTCAGTGCAGTACAGAGAGACA-----GTGAATATCC 1537
QY 441 ArgCysVal-----GluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGly 458
DB 1538 TTATGTACTAGAAAGAAATCTTAAGAGATGTCACATCTCTCTTATTAAACATGGA 1597
QY 459 AspThrThrSerPheProLeuSerValTyrProProGluSerThrValThrTyrArgCys 478
DB 1598 GTCATTTATTAATGTTCAACAGTACACCTTATGAATAGGCTCTTCACTAGATACAGATG 1657
QY 479 GlnSerPheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrSer 498

```

Db      1658 TTTGATCACCATTTCCTAGAGAGATCTAGGAGCCCTATTGTTTAAATGCAATGTCGACT 1717
QY      499 GIUPROFARAGCysLeuAspProCysValValSerGluGluAsnMetLysAsnAsn 518
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1718 ACACCAACCATGCTTTAGAGCATGACACTTATCTTTACTGAATGCAAAAGCAATAT 1777
QY      519 ILGlnLeuLysTrpArgAsnAspGlyLysLeuTrpAlaLysThrGlyAspAlaValGlu 538
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1778 TTACTTCTGAAATGAGATTTTGGCAATGAACCAATTTTTCATGCTGCAATATATTTGAG 1837
QY      539 PheGlnCysLys-----PheProHisLysAlaMetLysSerSerProPheArg 555
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1838 TTTATTTTGTAGAGACATCTTATCCAGCTGAATTTATATTACTGATCTATACTTACA 1897
QY      556 AlaIleCysGlnGluGlyLysPheGluTrpIleCys 568
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1898 ATGCAATGTGACAGAGCGCAATTAAATATCAAAAGATGT 1936

```

RESULT 12.

ABN97281
ID ABN97281 standard; DNA: 1269 BP.

AC ABN97281:

DF 13-AUG-2002 (first entry)

DE Gene #3779 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KM metastatic liver tumor; cytostatic; expression profile; disease state;

XX disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001: 2001MO-US30589.

XX 02-OCT-2000: 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX MPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,

XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,

XX PT involves detecting the level of expression of two or more genes in a

XX liver tissue sample

XX Claim 1: SEQ ID NO 3779: 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the

XX progression of liver cancer, hepatocellular carcinoma or metastatic liver

XX tumor in a patient, and differentiating metastatic liver cancer from

XX hepatocellular carcinoma in a patient, involving detecting the level of

XX expression of two or more genes represented in ABN93503-ABN97455 in a

XX tissue sample. The method of the invention has hepatotropic, and

XX cytostatic activity. The method is useful for diagnosing and detecting

XX the progression of liver cancer, hepatocellular carcinoma and metastatic

XX liver carcinoma in a patient. The method is useful for identifying

XX expression profiles which serve as useful diagnostic markers as well as

XX markers that can be used to monitor disease states, disease progression,

XX drug toxicity, drug efficacy and drug metabolism.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 1269 BP; 450 A; 224 C; 221 G; 374 T; 0 other;

Alignment Scores:

Pred. No.: 2, 2e-53 Length: 1269
Score: 710.00 Matches: 156
Percent Similarity: 37.72% Conservative: 59
Best Local Similarity: 22, 37% Mismatches: 115
Query Match: 22, 31% Indels: 240
DB: 24 Gaps: 8

US-09-805-337A-2 (1-569) x ABN97281 (1-1269)

```

QY      1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGlu 20
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      24 ATGTTCTTACTATTCATATCATGCTTACCTGTGCTTCTGCTTATGCAAGTG 83
QY      21 ThrLeuAspPheProLysIleHisIleGlyPheLeuTrpAspGluLysAspTyra 40
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      84 AAACCTTGATTTTTCAGACANTTAACATGAGAGCTATTTCATGACAAATTCCTACA 143
QY      41 ProPheSerGlnValProThrGlyGluValPheTyTrpSerCysGluTyraPheVal 60
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      144 CCATCTTCCAGTACTGCTAGCAAAATATATCTCTATTTACTGATGACATTTTGAG 203
QY      61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluLysTrpSerProThr 80
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      204 ACTCCGTGAGAGAGTACTGGGATTCATTCATTCGACCAAAATGCGGTGTCACGACA 263
QY      81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGlySer 100
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      264 GATCAGCATGCTCAGAAATGTTATTTCTTATTGGAATATGATATATCAAAATTAT 323
QY      101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSer 120
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      324 GGAGAAAGCTTGTGACAGGCTACTACAGAAAGTCCCTGCACTCTGCTACGCTCT 383
QY      121 GlnAsn---AsnGluLysAsnIleSerCysValGluArgLysTrpSerThrProProl 139
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      384 CCAAAAGTCCGTCAGCACACAGTTCATTCAGGAGAAATGCGTCTCTACTCCACA 443
QY      140 CysSerPheThrLysGlyLysValProIleLeuLysAsnValAspAlaGln 159
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      444 TGC----- 446
QY      160 ProLysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeu 179
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      447 -----ATCCGACAGACAAACATGCATCAAAATCCAGAT 476
QY      180 IleArgValGlySerAspSerValGlnCysTyrglnPheGlyTrpSerProAsnPhePro 199
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      477 ATA----- 479
QY      200 ThrCysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluVal 219
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      480 -----GAAATTCAAATGCAATTCAT 500
QY      220 LysGluIleArgLysGluLysValGlyHisAsnGluValGluTyraAspCysAsnPro 239
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      501 TCTGAA---TCTTCCCTATTATATATTAAATTAACAAATGTAACCA 557
QY      240 AsnPheIleIleAsnGlyProLysLysIleGlnCysValAsnGlyGluTrpThrLeu 259
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      558 GGNAT-----GCAACAGCAGATGCAAT----- 581
QY      260 ProThrCysValGluGlnValLysThrCysGlyTrpIleProGluLeuGluLysGlyTrp 279
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      581 ----- 581
QY      280 ValGlnProSerValProProTrpGlnHisGlyValSerValGluValAsnCysArgAsn 299
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      581 ----- 581
QY      300 GluTyraIleMetIleGlyAsnAsnMetIleThrCysIleAsnGlyTleTrpThrGluLeu 319
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

Db 582 -----TCTTCAGATCAATTAATGTTGCGA----- 608
OY 320 PrometCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLys 339
Db 608 ----- 608
OY 340 ThrIleuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCys 359
Db 608 ----- 608
OY 360 SerAspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGluVal 379
Db 609 -----AATGGA----- 614
OY 380 AspCysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGln 399
Db 614 ----- 614
OY 400 AsnMetThrThrValAsnTyrGlnAspGlyLysValAlaValLeuCysIysGln 419
Db 614 ----- 614
OY 420 AsnTyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrProIleSerLeu 439
Db 615 -----TGGTCAGACCA 626
OY 440 ProArgCysValGluSerThrAlaTyrCysGlyProProSerIleAsnAsnGlyAsp 459
Db 627 CCAATTGGATTAATTTCTTCAGAAAAGTGTGACCTCCCTCCACTTTAGCANTGGTGAT 686
OY 460 ThrThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGln 479
Db 687 ACCACCTCCTTCTACTAAAGTATATGCGCACAGTCAAGAGTCGAGTACCAATGCCAG 746
OY 480 SerPheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysIleThrPseGlu 499
Db 747 TCTACTATGAACCTTCAGGGTCTTATTATTGTAACATGTAATGAGAGTGTGCGCA 806
OY 500 ProProArgCysLeuAspProCysValValSerGluGlnAsnMetAsnLysAsnAsnIle 519
Db 807 CCACCTAGATGATCATCATCATGTAATTAATACGAAAGAAACATGATATAAATAATACATA 866
OY 520 GlnLeuLysTyrArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPhe 539
Db 867 AAGTTAAAGAAAGAAAGTACAGAAAATATTATGCCAAAAACAGGGGATACCATTTGAATT 926
OY 540 GlnCysLysPheProHisLysAlaMetIleSerSerProPheArgAlaIleCysGln 559
Db 927 ATGTGTAATTTGGATATATGCAAAATACATCAATTCTATTCATTCACAGAGTGTGCG 986
OY 560 GluGlyLysPheGluTyrProIleCysGln 569
Db 987 GAAGGATAGTGAATACCCAGATGCGAA 1016

```

RESULT 13
 AAS06029 standard; DNA; 1929 BP.
 AAS06029;

12-SEP-2001 (first entry)
 Angiotensin converting enzyme (ACEV) splice variant DNA #29.
 Angiotensin converting enzyme splice variant; ACEV: interleukin 6;
 granulocyte colony stimulating factor receptor; glucagon; hypertrophy;
 platelet-derived endothelial cell growth factor; cardiovascular disease;
 cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1c; ds;
 vasorelaxant intestinal polypeptide receptor 2; arteriosclerosis; cancer;
 myocardial infarction; coronary arterial thrombosis; renal disease;
 diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;
 multiple sclerosis; immune complex nephritis; deep vein thrombosis;
 nonatrophic pulmonary granulomatous disease; endothelial abnormality;

```

KM vascular disorder; asbestosis.
XX
OS Mus sp..
XX
PN WO200136632-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000: 2000MO-IL00766.
XX
PR 17-NOV-1999: 99IL-0132978.
PR 10-DEC-1999: 99IL-0133455.
XX
PA (COMP-) COMPUGEN LTD.
PI Levine Z, David A, Azar I, Khosravi R, Bernstein J:
DR MPI: 2001-336004/35.
DR P-PSDB: AAU02929.
XX
PT Novel alternative splicing variants e.g. variant of angiotensin
PT converting enzyme (ACEV), useful in identifying candidate compounds
PT capable of binding to the variant and to detect anti-variant antibodies
PT
PS Claim 1: Page 330-331; 519pp; English.
XX
CC The sequence represents a DNA encoding an angiotensin converting enzyme
CC splice variant (ACEV) polypeptide. The polypeptides of the invention
CC include variants of granulocyte colony stimulating factor receptor,
CC glucagon, interleukin 6, platelet-derived endothelial cell growth factor,
CC cyclin-dependent kinase inhibitor 1c, cellular tumour antigen p53, and
CC vasorelaxant intestinal polypeptide receptor 2. The polypeptides and their
CC associated nucleic acids are useful for identification of variant
CC sequences and detection of candidate compounds capable of binding the
CC molecules. The sequences of the invention can be used in the treatment
CC and diagnosis of various disorders including cardiovascular diseases such
CC as arteriosclerosis, myocardial infarction and coronary arterial
CC thrombosis, renal diseases such as diabetic nephropathy, muscular
CC diseases such as hypertrophy, immune disorders such as immune complex
CC nephritis, multiple sclerosis, cancer, sarcoidosis, nonatrophic
CC pulmonary granulomatous diseases such as asbestosis and vascular
CC pathologies involving an endothelial abnormality such as deep vein
CC thrombosis.
XX
SQ Sequence 1929 BP: 605 A; 385 C; 460 G; 479 T; 0 other:

```

Alignment Scores:
 Pred. No.: 1,34e-50 Length: 1929
 Score: 681.50 Matches: 160
 Percent Similarity: 36.77% Conservative: 72
 Best Local Similarity: 25.36% Mismatches: 206
 Query Match: 21.41% Indels: 194
 DB: 22 Gaps: 9

US-09-805-337A-2 (1-569) x AAS06029 (1-1929)

```

OY 7 ValIleLeuIleSerTyrPalaSerThrValGlyGluGlyThrLeuCysAspPhePro 26
Db 63 ATTCTTTATTAATCTCTTCAGAGAACTGTATGCAAGAGAAACAGTGTGATTTTCTCT 122
OY 27 LysIleHisIscIly-----PheLeuTyrAsp---GluLysAspTyrAsnPro 41
Db 123 ACCGTGGAATAATGAGAGATGAGTGGCCCATATTATATACGTTTAAAGCTTTACTCCCC 182
OY 42 PheSerGlnValProThrGlyGluValPheTyrTyrSerCysGlnTyrAsnPheValSer 61
Db 183 ATGAGC-----GTAGACAAAAACATATCATCTCTCTGTTGGCTGCTATGCAACC 233
OY 62 ProSerLysSerPheThrPheArgIleThrCysThrGluGluGlyTyrPseThrPro 81
Db 234 GAAGTGGGAGAGCAAGAGAGCAATTCAGTGCAGCAGAGAGGCTGCTCCAAACCA 293

```

```

QY 82 LysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGly 101
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 294 AGGTGCTACAGAAATGCTGCAAGCTTCAAGATGCTTCCATGCAAA 353
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 102 LeuIleHisLeuGluLysPheValGlnIleIleCysAsnThrGlyYserLeuGln 121
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 354 GATCTGTCAAACTTCAGAGAGATGACCTACGCTGCTTCAGATCAAAACCACT 413
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 122 -----AsnAsnGlyLysAsnIleSerCysValGluArgGlyTrpSerThrProPhe 139
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 414 GGGGGAAGAGAGAGAGAGGTGGTCCACCTCTCTTCGCAAGGAGGTCTTCAACGCTCC 473
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 140 CysSerPheThrLysGlyLysCysHisValProIleLeuGln----- 153
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 474 TGCAGAAAGAAACAGAGACATGTTGGCCCTGAGTGGAAACATGAATTAATCTCAGC 533
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 153 ----- 153
Db 534 ACGCAGAGAACTTCAAACTGAAGACATAGTGGCTTACACATGCACTGCAGGCTACTAC 593
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 153 ----- 153
Db 594 ACCACTACAGGAGAGACAGCTGGGGAAGCAGAGTGCAGGCTAATGGCTGGTCCCTCACA 653
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 154 -----AlaAsnValAspAlaGln 159
Db 654 CCACAGTCCCAAAATTAATGTGCTCTTCTTGCAGATTAATAGAAATGGCTATTTCAT 713
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 160 ProLysLysGlySerThrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeu 179
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 714 CCTGTAAAGCAACATATGACAGAGAGATGATGATCAATTTTTCGTATGAAATATAT 773
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 180 IleArgValGlySerAspSerValGlnCysTyrglnPheGlyTrpSerProAsnPhePro 199
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 774 TATCTAAGAGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 833
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 200 ThrCysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyVal 219
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 834 ATATGTGAAGAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 890
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 220 LysGluIleArgLysGluLutrgLysHisAsnGluValGluTrpAspCysAsnPro 239
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 890 ----- 890
QY 240 AsnPheIleIleAsnGlyProLysLysIleGlnCysValAspGlyLutPrThrThrLeu 259
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 890 ----- 890
QY 260 ProThrCysValGluGlnValLysThrCysGlyTrpIleProGlnLeuGlnLutrgLys 279
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 890 ----- 890
QY 280 ValGlnProSerValProProTyrglnHisGlyValSerValGluValAsnCysArgAsn 299
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 891 ATTCACACCAATTCACCAATTCACCAATTCACCAATTCACCAATTCACCAATTCACCAATTC 950
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 300 GluTyrglnMetIleGlyAsnAsnMetIleThrCysIleAsnGlyLutPrThrThrLeu 319
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 951 AA-TTTGTGATTCACAGGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1009
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 320 PrometCysValAlaIleThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLys 339
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1010 CCAAAATGATTT----- 1021
QY 340 ThrLeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrgCys 359
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1021 ----- 1021
QY 360 SerAspIlePheArgTyrgArgHisSerValCysIleAsnGlyLysTrpAsnProGluVal 379
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1021 ----- 1021
QY 380 AspCysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGln 399

```

```

Db 1022 -----GAGGAAGAGAGAGAGTACCTGCGAGCAACCCCTCCGTTGAGCAATGCTTT 1075
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 400 AsnMetThrThrValAsnTyrglnAspGlyGluLysValAlaValLeuLysGlu 419
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1076 GCACACCCACACTGTGATTTACTACTGAGGAGTAAAGATACATACAGGTGTGAGCG 1135
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 420 AsnTyrgLeuProGluValAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeu 439
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1136 GGCCTACAGCTCAGAGAGATCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 1195
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 440 ProArgCysValGluSerThrAlaTyrgCysGlyProProProSerIleAsnAsnLysp 459
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1196 CCTGAGTGTGTAAGATTTGAGATTTGAGATTTGAGATTTGAGATTTGAGATTTGAGATTT 1255
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 460 ThrThrSerPheProLeuSerValTyrgProGlySerThrValThrTyrgCysGln 479
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1256 GTTGTGATGGCTTACTGGCAAGTTACACACAGATCTTCTGTGAGTACCGATGCAAT 1315
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 480 SerPheTyrgLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGln 499
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1316 GAGTATTACTTATTGAGAGGTCGGAACATCTCCCTGTGAACAAGACCGTGGTCACT 1375
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 500 ProProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnIle 519
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1376 CCACCTGCTGCTGAGAGCCCTGACATATGATGATGATGATGATGATGATGATGATGAT 1435
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 520 GlnLeuLysTrpArgAsnAspGlyLysLeuTyrgAlaLysThrGlyAspAlaValGluPhe 539
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1436 CAGTTGAGTGAATATGACGAGGCAAGATTTTA-----CATGAGACCTTAATGATTTT 1489
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 540 GlnCysLysPheProHisLysAlaMetIleSerSerPro-----PhePheArgAlaIle 557
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1490 GTGTGTAACAGAGGATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1549
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 558 CysGlnGluGlyLysPheGluTyrgProIleCys 568
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1550 TGCATATAGAGGAGATGTGAGTACCCCATGTC 1582
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

RESULT 14
ABN97342
ID ABN97342 strand: DNA; 1315 BP.
AC ABN97342;
DT 13-AUG-2002 (first entry)
XX DE Gene #3840 used to diagnose liver cancer.
XX KW Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX KW metastatic liver tumour; cytostatic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX Homo sapiens.
XX WO200229103-A2.
XX 11-APR-2002.
XX PD 02-OCT-2001; 2001WO-US30589.
XX PF 02-OCT-2000; 2000US-237054P.
XX PR 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX PA Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX PI WPI: 2002-426119/45.
XX DR
XX PT Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a
XX PT liver tissue sample


```

RESULT 15
AAV02792
ID AAV02792 standard: DNA: 767 BP.
XX
XX AAV02792:
AC
XX
XX 27-APR-1998 (first entry)
XX
XX Human partial complement factor H cDNA fragment 2.
DE
XX
XX Complement factor H; tumour associated antigen; renal cancer;
KM urogenital cancer; medicament; modulator; ss.
XX
XX Homo sapiens.
OS
XX
XX MO9738136-A1.
PN
XX
XX 16-OCT-1997.
PD
XX
XX 09-APR-1997: 97MO-US05710.
PF
XX
XX 06-MAR-1997: 97US-0812481.
PR
XX 09-APR-1996: 96US-0015083.
PR 09-APR-1996: 96US-0630048.
PR 06-MAR-1997: 97US-0038614.
XX
XX (BARD-) BARD DIAGNOSTIC SCI INC.
PA
XX
XX Enfield DL, Hass GM, Kinders RJ;
PI
XX
XX WPI: 1997-512742/47.
DR
XX
XX P-PSDB: AAW39156.
XX
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement factor
XX H related antigen, or nucleic acid encoding it
XX
XX Example 6B; Fig 7A: 104pp: English.
XX
XX This partial cDNA sequence encodes a region of the human
XX tumour-associated complement factor H (CFH). This sequence is used
XX in the identification of DNA encoding complement factor H related
XX proteins and antigens from clone PBBSF2910 (see AAV02793-V02795). The
XX detection of this protein and a CFH antigen can be used in screening or
XX for the treatment of renal or urogenital cancer, e.g. bladder, cervical
XX or prostate cancer. Agents that may modulate this antigen could be used
XX in the manufacture of a medicament for the treatment of a tumour cell.
XX
XX
XX Sequence 767 BP; 257 A; 140 C; 179 G; 191 T; 0 other:
SO
Alignment Scores:
Pred. No.: 2,25e-36 Length: 767
Score: 515.00 Matches: 105
Percent Similarity: 47.47% Conservative: 45
Best Local Similarity: 33.23% Mismatches: 90
Query Match: 16.18% Indels: 76
DB: 18 Caps: 6
US-09-805-337A-2 (1-569) x AAV02792 (1-767)
QY 225 GIUGLUTYRGYIHISASGLUVALYALGIUTYRASPCYSASNPPOASNPHELLEIASN 244
DB 25 GACAGTTACAGTATGAGAGAGAGAGTACGTACCAATGTTTGAAGTTCGAAATGTGAT 84
QY 245 GLYPTOLYSLSIILEGLINCYSVALASPGIYUPTPTHTHLEUPROTHRCYSVALG 264
DB 85 GGGCCCTGCATTTGCCAAAATGCTTAGAGAGAAAATGCTCACCCTCCATCATGATAA 144
QY 265 GINVALYSTHCYSGLYTYRIILEPROGLIENGLIUTYRGYTYRVALG 284
DB 145 -----ACGATGTGCTCTAGTTACCTAGCTTGAA-----AATGCCATA 183

```

```

QY 285 PRO-----ProTYRGINHSIGLYVALSERVALGLUVALASNCYSARG 298
DB 184 CCCATGGGAGAGAGAGAGATGTGTATAGGGCGGAGACAGTACTTACACTTGTGCA 243
QY 299 ASNGIUTYRALMETLEGLYASNSMETLETHRCYSIILEASNGIYLETPTHTG 318
DB 244 ACATATTACAAAATGATGAGCCAGTAATGATACATTAATAGCAGATGACAGCA 303
QY 319 LEUPROMETCYSVALALATHRHISGLINLEULYSARGCYSYLIILEAGLYVALIS 338
DB 304 AGCCCAACATGC----- 315
QY 339 LYSTHLEULEULYSLEUSERGILYSGIUPHEASNHISASNSERARGILEARGTYRARG 358
DB 315 ----- 315
QY 359 CYSSEASPILLEPHEARGTYRARGHISSEVALCYSIILEASNGIYLETPTHTG 378
DB 316 -----AGAGACACCTCTCTGTGTAAT----- 336
QY 379 VALASPCYSTHRLULYSARGLUGLINPHECYSPTROPTROPROGLINLEPROASNA 398
DB 337 -----CCGCCACAGTACAAATGCT 357
QY 399 GLINAMETHTHTHTRVALASN---TYRGLINASPGIYGLULYSVALALAVALL 417
DB 358 TATATAGTCTCGACAGACAGATGATTAATCCATCTGTGAGACAGATGATCAATCT 417
QY 418 LYSGLINASTYRLEULEUPROGLUALALYSGLIUILEVALCYSGLYASPGIYARGT 437
DB 418 AGGAGCCCTTATGAATGTTTGGGATGAGAGATGATGTGTTAAATGGAAGACTGGACG 477
QY 438 SERLEUPROARGCYSVALGUISERTHTRALATYRCYSGIYPTROPTROSERI 457
DB 478 GAACCACTCATATGCAAAATGCTTACAGAAATGAGGGCCCTCCACCTATGACAAT 537
QY 458 GLYASPTHTHSETPHEPROLEUSERVALITYRPTROPTROGLYSETHTYRVAL 477
DB 538 GGGGACATTACTCATTCCTCCGTTGCAGTATATGCTCCAGCTTCAGATTGACAGCAA 597
QY 478 CYSGLINSEPTHTYRLEULINLEULINLEUSERVALITHRVALITHRCYARGAS 497
DB 598 TCCCAAGACTTATCAACTTACAGGTAACAAGCAATATACATGTAGAAATGACAAATG 657
QY 498 SERGLIUPROARGCYSLEUASPTROCYVALVALISERGLUGLUASNMETAS 517
DB 658 TCAGAACCAACCAAAATGCTTACATCCGTTGTAATATCCCGGAATATATGAAATAT 717
QY 518 ASNILEGLINLEULYSTPATRGASNASPGIYLSLEUTYRALALYSTH 533
DB 718 AACATAGCATTTAAGGTGAGACAGCAACAGAGAGCTTTATTCGAGACA 765

```

Search completed: July 23, 2003, 09:03:04
Job time : 368 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2003, 03:08:43 ; Search time 2198 Seconds
(without alignments)
4192.553 Million cell updates/sec

Title: US-09-805-337a-2
Perfect score: 3183
Sequence: 1 MLTFSVILISWTVGEG.....SSPPRALCOEGKFEYPICE 569

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV-rlh
-O/cgn2.1/uspro.spool/us09805337/runat.17072003.135605.18928/app_query.fasta.1.711
-DB-est -QFMT-fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=100 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09805337 @cgn2.1.1906 @runat.17072003.135605.18928 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -NARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estinu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gD_est1:*
10: gD_est2:*
11: gD_hic:*
12: gD_est3:*
13: gD_est4:*
14: gD_est5:*
15: em_estfun:*
16: em_estom:*
17: gD_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2384	74.9	1411	11	BC026282	BC026282 Homo sapi
2	1037	32.6	683	12	BG616762	BG616762 602614967
3	1016	31.9	646	10	AV657525	AV657525 AV657525
4	1014	31.9	633	10	AV657514	AV657514 AV657514
5	828.5	26.0	795	10	AV683459	AV683459 AV683459
6	801	25.2	832	12	BG562869	BG562869 602581712
7	797.5	25.1	858	12	BG569423	BG569423 602586766
8	792.5	24.9	853	9	AU138301	AU138301 AU138301
9	788.5	24.8	912	12	BG564962	BG564962 602583932
10	785.5	24.7	786	12	BG533258	BG533258 601860623
11	780	24.5	839	12	BG533248	BG533248 601860612
12	777	24.4	790	12	BG286384	BG286384 602383156
13	777	24.4	815	10	AV682232	AV682232 AV682232
14	764	24.0	761	10	AV688058	AV688058 AV688058
15	760.5	23.9	735	12	BG616866	BG616866 602615818
16	759.5	23.9	787	12	BG563052	BG563052 602581839
17	759	23.8	693	10	AV649114	AV649114 AV649114
18	758	23.8	723	10	AV682522	AV682522 AV682522
19	757.5	23.8	846	12	BG616668	BG616668 602614870
20	757	23.8	739	12	BG563738	BG563738 602584521
21	753	23.7	801	12	BG563289	BG563289 602582225
22	749	23.5	675	13	B1254780	B1254780 602974005
23	745.5	23.4	823	12	BG563035	BG563035 602581819
24	742.5	23.3	777	12	BG567040	BG567040 602589358
25	739.5	23.2	838	12	BG617334	BG617334 602615382
26	735	23.1	721	12	BG568304	BG568304 60267230
27	727.5	22.9	650	12	BG565462	BG565462 602583417
28	726	22.8	744	12	BG563564	BG563564 602581978
29	724.5	22.7	739	9	AU139391	AU139391 AU139391
30	721.5	22.7	681	12	BG568358	BG568358 602587292
31	721.5	22.7	780	12	BG565469	BG565469 602583426
32	720	22.6	1113	12	BG567998	BG567998 602586857
33	719	22.6	668	10	AV650989	AV650989 AV650989
34	718.5	22.6	739	12	BG568871	BG568871 602616154
35	716.5	22.5	886	12	BG568871	BG568871 602588041
36	715.5	22.5	771	10	AV682102	AV682102 AV682102
37	715.5	22.5	777	12	BG618219	BG618219 602645945
38	714	22.4	785	12	BG569316	BG569316 602588621
39	714	22.4	800	12	BG568888	BG568888 602588060
40	713.5	22.4	836	12	BG564780	BG564780 602584012
41	708.5	22.3	734	12	BG617992	BG617992 602644866
42	708.5	22.3	880	12	BG569198	BG569198 602588472
43	705	22.1	643	10	AV681962	AV681962 AV681962
44	703.5	22.1	802	12	BG617178	BG617178 602615575
45	701.5	22.0	617	12	BG563181	BG563181 602582092

ALIGNMENTS

RESULT 1
BC026282
LOCUS
DEFINITION
Homo sapiens, similar to factor H-related protein 5, clone
IMAGE:4733780, MRNA.
ACCESSION
BC026282
VERSION
BC026282.1 GI:20070808
KEYWORDS
HTC.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1411)
AUTHORS
Strausberg, R.
TITLE
Direct Submission

JOURNAL

Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdexaxl@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: RAL Plate: 37 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13540562
This clone has the following problem: frame shifted.

FEATURES

Source

1. 1411
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4733780"
/tissue_type="Liver"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

BASE COUNT

496 a 261 c 275 g 379 t

Alignment Scores:

Pred. No.: 6 45e-247 Length: 1411
Score: 2384.00 Matches: 427
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 74.90% Indels: 0
DB: 11 Gaps: 0

US-09-805-337a-2 (1-569) x BC026282 (1-1411)

QY 1 MetLeuLeuLeuPheSerVal11LeuLeuIleSerTrpValSerThrValGlyGlyGly 20
DB 110 ATGTTCTCTTATTCAGTGAATCTAATCTCATGCTATCCACTGTGGGAGAGAGA 169
QY 21 ThrLeuGlyAspPheProLysIleHisGlyPheLeuTrpAspGluLysPtyrAsn 40
DB 170 ACACCTTTGATTTTCCAAAATAACACCATGATTTCTGATGATGAGAGATTATAAC 229
QY 41 ProPheSerGlnValProThrGlyGluValPheTrpTrpSerCysGluLysAsnPheVal 60
DB 230 CCTTTTCCCAAGTCTCTACAGGAGGAGTTTCTATTACTCTCGTAATATTAATTCG 289
QY 61 SerProSerLysSerPheTrpPheTrpArgIleThrCysThrGluGluGlyTrpSerProThr 80
DB 290 TCTCCTTCAAAATCCTTTTGAGACTCGCATACATGACAGAGAAGATGTCACCAACA 349
QY 81 ProLysGlyLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer 100
DB 350 CCGAAGTGTCTCAGATGTCTTCTTCTTTGTAAGAAAGGTCATTCGAATCTTCA 409
QY 101 GlyLeuIleHisLeuGluLysAspPheValGlnIleIleCysAsnThrGlyTrpSerLeu 120
DB 410 GGACTAATATCATCTGGACGCTATCTGACAAATTAATTGACACAGATACAGCTT 469
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProPheLys 140
DB 470 CAAACAAATGACAAAAACATTTCTGTGTAGAACGGGCTGTGCTCCACTCTCCATATGC 529

QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluValAsnValAspAlaGlnPro 160
DB 530 AGCTTCATTAAGGACAAATGTATGTCCAAATTTTGAAGCCAAATGATATCTCAGCCA 589
QY 161 LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
DB 590 AAAAAAGAAAGCTACAAAGTTGGACACGTTGCAAAATCTCTCCGCAAAAAATCTTATA 649
QY 181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr 200
DB 650 AGAGTTGGATCAGACACCACTTCAATGTTCACAAATTTGGTGTGCTACCTTATCCACA 709
QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyLysValLys 220
DB 710 TCCAAAGCAGACAGTACATCATGTGTGCTCCACTCTCCAACTCCCAATGTGTAAGTAA 769
QY 221 GlnIleArgLysGluLysTrpGlyHisAsnGluValGluTrpAspCysAsnProAsn 240
DB 770 GAGATAGAAAGAGAGAAATATGACACAAATGAGTGGAAATATGATTCATCTAAT 829
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260
DB 830 TTTATATATAACGGCGCTTAAGAAATACATGTGTGATGAGCAATGACAACTTACC 889
QY 261 ThrCysValGluGluValLysThrCysGlyTrpIleProGluLeuGluLysTrpVal 280
DB 890 ACTTGCTTGAACAAAGTGAACAAATGTGATACATCTGACACTGACAGTACGTTATGTT 949
QY 281 GlnProSerValProProTrpGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
DB 950 CAGCCCTGTCCCTCCCTCATACACATGAGATTTCACTGACGTGAATGCAAAATGAA 1009
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320
DB 1010 TATGCAATGATTTGAAATATACATGATTAATCTGATTAATGATATGACAGACTCTCT 1069
QY 321 MetCysValAlaThrHisGlnLeuLysAspGlyLysIleAlaGlyLysAsnIleLysThr 340
DB 1070 ATGTGTGTGGCAACACCACTTAAGAGCTGCAAAATAGCAGAGATTAAATTAAGAA 1129
QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTrpArgCysSer 360
DB 1130 TTACTCACTTATCTGGGAGAGAAATTTAATCAATTAATCTAGAAATGCTTACAGATGTCA 1189
QY 361 AspIlePheArgTrpArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
DB 1190 GACATCTTCAGATACAGGACACTGCTGTATTAACGGGAATGCAATCTCAGATAGAC 1249
QY 381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
DB 1250 TGCACAGAAAAAGGAAACAAATTCGCCACCGCACCTCAGATACCTAATCTCAGAA 1309
QY 401 MetThrTrpThrValAsnTrpGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
DB 1310 ATGACACACCACTGATATATCAAGGTGAGAAAAAGTAGCTGTCTGTATGAAGAAAC 1369
QY 421 TyrLeuLeuProGluAlaLysGlu 428
DB 1370 TATCTACTTCCAGAAAGCAAAAAA 1393
RESULT 2
Bg616762 683 bp mRNA linear EST 18-APR-2001
LOCUS BG616762 602614987FL NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4733780 5',
DEFINITION mRNA sequence.
ACCESSION BG616762
VERSION BG616762.1 GI:13668133
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 683)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 DNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1CM1595 row: n column: 21
 High quality sequence stop: 683.

FEATURES
 source Location/Qualifiers
 1.683

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH-MGC-76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccatagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATATAGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 220 a 132 c 130 g 201 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,35e-101 Length: 683
 Score: 1037.00 Matches: 190
 Percent Similarity: 99.48% Conservative: 0
 Best Local Similarity: 99.48% Mismatches: 0
 Query Match: 32.58% Indels: 1
 DB: 12 Gaps: 0

US-09-805-337A-2 (1-569) x BG616762 (1-683)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
 |||||
 Db 111 ATGTTGCTCTTATTCAGTGAATCCCTAATCTCATGGTATCCACTGTTGGGGAGAGGA 170
 QY 21 ThrLeuCysAspPheProIysIleHisIleSgIyPheLeuTyraSpGluGluAspTyra 40
 |||||
 Db 171 ACACCTTGATGATTTCCAAAATAACACCATGATGATTTCTGATGATGAAGATTTAATAC 230
 QY 41 ProPheSerGlnValProThrGlyGluValAlaPheTyrTyrSerCysGluTyrAsnPheVal 60
 |||||
 Db 231 CCTTTTCCCAAGTCTTACAGGGGAGGATTTCTATTACCCCTGTAATTAATTTTCGTG 290
 QY 61 SerProSerIysSerPheTrpPheArgIleThrCysThrGluGlyTyrSerProThr 80
 |||||
 Db 291 TGTCCCTTCAAAATCCCTTGGACTCCCATACATGCACAGAAAGAGATGTCACCAACA 350
 QY 81 ProIysCysLeuArgMetCysSerPheProPheValIysAsnGlyHisSerGlnSerSer 100
 |||||
 Db 351 CCGAATGCTCTCAGATGCTGCTTCTTCTTGTGAAAAATGCTCATTTCTGAATCTTCA 410
 QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleLeuCysAsnThrGlyTyrSerLeu 120
 |||||
 Db 411 GGCCTAATACATCTGAGAGCTGATCTGACAAATATTTGCAACACAGATACACGCTT 470
 QY 121 GlnAsnAsnGluIysAsnIleSerCysValGluArgGlyTyrSerThrProPheIleCys 140
 |||||
 Db 471 CAAACAAATGAGAAAAACATTTGTGTGTAACAGGGGCTGCTCCACTCTCCCATATGC 530
 QY 141 SerPheThrIysGlyIleCysHisValProIleLeuGluValAsnValAspAlaGlnPro 160

|||||
 Db 531 ACCTGACATTAAGAGAGATGTATGTTCCAAATTTTGAAGCCATGATGCTACGCCA 590
 QY 161 LysLysGlnSerTrpIlyValGlyAspValLeuIlyPheSerCysArgIysAsnLeuIle 180
 |||||
 Db 591 AAAAAAGAAAGCTTACCAAGTGGAGACGTTGTAAATTTCTCCGACGAAAAAATCTTATA 650
 QY 181 ArgValGlySerAspSerValGlnCysTyr 190
 |||||
 Db 651 CAGAGTTGATTCAGACTCATGTTTAC 681

RESULT 3
 AV657525 646 bp mRNA linear EST 16-JAN-2002
 LOCUS AV657525
 DEFINITION AV657525 GLC Homo sapiens cDNA clone GLCFG11 3', mRNA sequence.
 ACCESSION AV657525
 VERSION AV657525.1 GI:9878539
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 646)
 Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Lu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Ou,J., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.

Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 21625106

JOURNAL MEDLINE
 COMMENT Contact: Zengqiang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source Location/Qualifiers
 1.646
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GLCFG11"
 /clone_lib="GLC"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
 Location/Qualifiers

BASE COUNT 201 a 132 c 121 g 190 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.32e-99 Length: 646
 Score: 1016.00 Matches: 188
 Percent Similarity: 87.56% Conservative: 2
 Best Local Similarity: 86.64% Mismatches: 8
 Query Match: 31.92% Indels: 19
 DB: 10 Gaps: 1

US-09-805-337A-2 (1-569) x AV657525 (1-646)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
 |||||
 Db 46 ATGTTGCTTATTCAGTGAATCCCTAATCTCATGGTATCCACTGTTGGGGAGAGGA 105
 QY 21 ThrLeuCysAspPheProIysIleHisIleSgIyPheLeuTyraSpGluGluAspTyra 40
 |||||
 Db 106 ACACCTTGATATTTTCCAAAATAACACCATGAGATTTCTGATGATGAAGAGATTTAATAC 165
 QY 41 ProPheSerGlnValProThrGlyGluValAlaPheTyrTyrSerCysGluTyrAsnPheVal 60

Db		166	CCTTTTCCCAAGTTCCTCAGACGGGAAGTTTCTATTACTCCTCGGATATTAATTTGTG	225
Oy		61	SerProSerIysSerPheTrpThrArgIleIleHicYstHrGlUGlyTYTrSerProthr	80
Db		226	TCTCCTCTAAATAATCCTTTTGACTCGCATTAACAATCACAGAAGAGVGTCCACAACA	285
Oy		81	ProIysCysLeuArqMetCysSerPheProHeValIlysAsnGLYHisSerGIUserSer	100
Db		286	CCGAAGTGTCTCAGANTGTGTCTCTTCCTTTGTGGAAAATGTGCATTCGAACTTCCA	345
Oy		101	GlyLeuIleHisLeuGlUGlyAsPThrValIleIleIleCysAsnThrGlyTYTrSerLeu	120
Db		346	GGACTTAATACATCTCGAAGCTGGAATACGTGTACAAATTAATTTGCACACAGATACAGCCTT	405
Oy		121	GLAsnaSnaGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys	140
Db		406	CAAAACAATGAGAAAAACATTTCGTGTGTACACCGNCTGTGCACCTCCCTCCCATATGC	465
Oy		141	SerPheThrIysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro	160
Db		466	AGCTTCACTAAGACAGAAATCTGCATGTTCCACATTTAGAACGCCAAATGATGATCAGCCACA	525
Oy		161	LysIysGluSerTyrlsValGlyAspAlleuLysPheSerCysArg-LysAsnLeuIl	180
Db		526	AAAAAAGAAACCTCACAGTTCGAGACCTGTGAAATTCCTCTCGACAGATCA-----	577
Oy		180	eArGVaIGlySerIsPserValGInCysTYrGlnPheGlyTrpSerProAsnPheProTh	200
Db		578	-----CCTAACTTTTCAAC	591
Oy		200	rCyAlsySGlyGlnValArqSerCysGlyProProProGlnLeuSerAsn	216
Db		592	ATGCANAGACAACTACGACATCATGTGTGCACCTTCCTCAATCTTCAAT	640
RESULT 4				
LOCUS	AV657514	633 bp	mRNA	linear EST 16-JAN-2002
DEFINITION	AV657514 GLC Homo sapiens cDNA clone GLOCFC12 3', mRNA sequence.			
ACCESSION	AV657514			
VERSION	AV657514.1	GI:9878528		
KEYWORDS	EST.			
SOURCE	Homo sapiens			
ORGANISM	human.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
	1 (bases 1 to 633)			
	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,			
	Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,			
	Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,			
	Hu,G., Gu,J., Chen,Z. and Han,Z.			
	Insight into hepatocellular carcinogenesis at transcriptome level			
	by comparing gene expression profiles of hepatocellular carcinoma			
	with those of corresponding noncancerous liver			
	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)			
TITLE	P16C25106			
JOURNAL MEDLINE	Contact: Zeguang Han			
COMMENT	Chinese National Human Genome Center at Shanghai			
	351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai			
	201203, P. R. China			
	Tel.: 86-21-50801919(ex. 45)			
	Fax: 86-21-50801922			
	Email: hanzg@chgc.sh.cn			
	This clone is available at CHOC in Shanghai.			
FEATURES	Location/Qualifiers			
Source	1..633			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="GLOCFC12"			
	/clone.lib="GLC"			
	/tissue_type="corresponding non cancerous liver tissue"			
	/dev_stage="Adult"			
	/lab_host="SOLR"			

BASE COUNT	193 a	132 c	123 g	185 t
ORIGIN	.			
Alignment Scores:				
Pred. No.:	3.7e-99	Length:	633	
Score:	1014.00	Matches:	186	
Percent Similarity:	87.85%	Conservative:	2	
Best Local Similarity:	86.92%	Mismatches:	8	
Query Match:	31.86%	Indels:	18	
DB:	10	Gaps:	1	
US-09-805-337A-2 (1-569) x AV657514 (1-633)				
Oy	1 MetleuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValIGlyGly 20			
Db	46 ATGTTGCTCTTATTCAGTGAATTCCTAATTCATATGGATATCCACTGTTGGGGACAGCA 105			
Oy	21 ThrIeuCysAspPheProLysIleHisIleGlyPheLeuTrpAspGluGluAspTrpAsn 40			
Db	106 ACACCTTGTGATTTCACAAATATACACCATGGAATTCGTATGATGAGACAGATTAATAC 165			
Oy	41 ProPheSerGluValProThrGlyGluValPheTrpTrpSerCysGluTrpAsnPheVal 60			
Db	166 CTTTTTTCACAAAGTCTCACAGGGGAAATTTTCTATTAATCTCTGGAATTAATTTTGTG 225			
Oy	61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr 80			
Db	226 TCTCTTCAAAATCTCTTTGGACCTGCGATACATGACACAGAAAGATGCTCACCAACA 285			
Oy	81 ProLysCysLeuAlaGlyMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100			
Db	286 CCGAAGTGTCTCAGAAATGTCTTCTCTTCTTGTGAAATATGGCATTCGAATCTTCA 345			
Oy	101 GlyIleuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu 120			
Db	346 GGACTAATACATCTCGAAGGTGATACCTTACAAATTAATTTCCACACACGATACAGCCTT 405			
Oy	121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys 140			
Db	406 CAAAACATATGAGAAATACATTTCTGTGTACACGGGGCTGTCCACCTCCCATATATGC 465			
Oy	141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160			
Db	466 AGCGTTACTTAAGGGAATGTCATGTTCCAAATTTTAGAGCAAGCAATGATGATGCTCAGACA 525			
Oy	161 LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180			
Db	526 AAACAAGACAGCTACACAGATGAGAGACTGTGAATTCCTCCTCGACA----- 573			
Oy	181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr 200			
Db	574 -----ACACCTTAATCTTTTCACCA 591			
Oy	201 CysLysGlyGlnValArgSerCysGlyProProGlnLeu 214			
Db	592 TGCACAGGACACAGTACGATCATGTGGGACCACCTCTTCACATT 633			
RESULT 5				
AV683459	795 bp mRNA linear EST 16-JAN-2002Z			
LOCUS				
DEFINITION	AV683459 GKC Homo sapiens cDNA clone GKCDP09 5', mRNA sequence.			
ACCESSION	AV683459			
VERSION	AV683459.1 GI:10285322			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Xiao,H., Xu,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xie,X., Huang,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,			

Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z. and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level

by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21635106

21023100 MEDLINE
CONTACT: Zeguang Han

Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel.: 86-21-50801919(ex.45)
Fax: 86-21-50801923

Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

source	location/yudalliers
1. .795	

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

```
/clone="GKCDPF09"  
/clone lib="GKC"
```

```

/tissue_type="hepatocellular carcinoma"
/row_stage="Stage1"

```

```
/dev_stage=Adult
/lab_host="SOLR"
```

```

/note="Vector: pbluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

```

BASE COUNT	254 a	160 c	151 g	227 t	3 others
ORIGIN					

Instrument: GABQ-10

Alignment scores:
red. No.: 6.22e-79 Length: 795

core:	828.50	Matches:	179
percent Similarity:	36.92%	Conservative:	20

Test Local Similarity:	33.21%	Mismatches:	39
Very Match:	26.03%	Indels:	302

B:	10	2
Gaps:		2

S-09-805-337A-2 (1-569) x AV683459 (1-795)

1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20

63 ATGTGGCTCGTGCAGTCAATTCTAATCTACGGATATCTGTGGGGGACAAGCA 122

21 **Phyllosticta dendrobii** (Uchiy. & Nishik.) Uehel. & Uchiy. *Ann. Bot. Soc. Japan* 40: 103 (1974)

```

21  in deucysprnerioulislenisnisiylnebeuiylaspoiuciuaspiylasn 400
      |||||:::||||| |||||

```

RESULT 6	BC562869	832 bp	mRNA	linear	EST 10-APR-2001
LOCUS	BC562869				
DEFINITION	602581712p1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4709426 5', mRNA sequence.				
ACCESSION	BC562869				
VERSION	BC562869.1	GI:13570521			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 832)				
TITLE	NIH-MGC http://mgc.ncl.nih.gov/ .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				

FEATURES	Location/Qualifiers
source	1. .832

BASE COUNT	275 a	171 c	154 g	232 t
ORIGIN				

Alignment Scores:

Pred. No.:	6,31e-76	length:	832
Score:	801.00	Matches:	179
Percent Similarity:	36.83%	Conservative:	21
Best local Similarity:	32.97%	Mismatches:	39
Query Match:	25.16%	Indels:	304
DB:	12	Gaps:	3

US-09-805-337A-2 (1-569) x BG562869 (1-832)

OY	MetLeuLeuPheSerValIleLeuIleSerTPaIleSerThrValGlyGlyCly	20
Db	83 ATGTGGCTCTGTGCTAGTAAATTCTTAACTTCACGGATATCTCTGTGGGGGAAAGCA	142
OY	21 ThrLeuCysAspPheProCysIleHisHisGlyPheLeuTyPaspGluAspTyPasn	40
Db	143 ATGTCTGTGATTTTCCAAAATAAACCATGGAAATTCTATATGATGAAGAATAAATTATAG	202
OY	41 ProPheSerGluValProThrGlyGluValPheTyTyPserCysGluTyPasnPheVal	60
Db	203 CCATTTTCCCAAGTTCCTACAGGGGAAGTTTCTATCTACTCGTGAATATAAATTTTGTG	262
OY	61 SerProSerLySerPheThrPheArgIleThrCysThrGluGluGlyTyPserProThr	80
Db	263 TCTCCCTTCAAAATCTTTTGGATCGCATAAAGTCCGCAAGAAGATGTCTCCACACA	322

QY	8	ProLysGluLeuArgHecTysSerPheProPheValLysAsnGluYHnSerGluSer	1000
Db	323	CCAAAGGCTCAGACTGCTGTTCTTCTCTTCTTGGAAGGATGGCATTCGAAATCTCA	3622
QY	101	GlyLeuLeuHisLeuGluGluGlyAspThrValGlnIleIleIleCysAsnThrGlyTrpSerLeu	1200
Db	383	GGACAAACACATCTGGAGGTGATCTGTACAAATATTCTGCACACAGATACAGACTT	4422
QY	121	GlnAsnAsnGluLysAsnIleSerCysValGluAtgGlyTrpSerThrProProIleCys	1400
Db	443	CAAAACATGTAGAACAACTTTCATGTGTAGAACGGGCTGTGTCACTCCCAAAATGC	5020
QY	141	SerPheThrLysGluLysLincysHisValProIleLeuGluAlaAsnValAspAlaGlnPro	1600
Db	502	-----502-----	
QY	161	LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle	1800
Db	502	-----502-----	
QY	181	ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr	2000
Db	502	-----502-----	
QY	201	CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLys	2200
Db	502	-----502-----	
QY	221	GluIleArgGlySerGluGluTrpGlyHisAsnGluValValGluTrpAspCysAsnProAsn	2400
Db	502	-----502-----	
QY	241	PheIleLeuAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro	2600
Db	502	-----502-----	
QY	261	ThrCysValGlnGlnValLysThrCysGlyTyrIleProGluLeuGluTrpGlyTyrVal	2800
Db	502	-----502-----	
QY	281	GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGlu	3000
Db	502	-----502-----	
QY	301	TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro	3200
Db	502	-----502-----	
QY	321	MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr	3400
Db	502	-----502-----	
QY	341	LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer	3600
Db	502	-----502-----	
QY	361	AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp	3800
Db	502	-----502-----	
QY	381	CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn	4000
Db	502	-----502-----	
QY	401	MetThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn	4200
Db	502	-----502-----	
QY	421	TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro	4400
Db	502	-----502-----	

Oy		441	ArgyrysalgiuserthrinalatrycysglYproproserilleasnannglyAspThr	460
Db		503	AGGCCACATTTTGTGCAGAAAAATGtGGCCCTCACCTATTGAACAATGAGACATT	562
Oy		461	ThrsrPheProLeuSerValTyProProglySerThrValThrTyArgCysGlnSer	480
Db		563	ACTTCATTCCTCGTTGTCACTATATAGCTCCAGAGTTCACTAGTACCATGCCAACAC	622
Oy		481	PheTrylrLysleugInglySerValThrValThrCysArgasnLysGlnTrpSerGluPro	500
Db		623	TTCATCAACTGTGAGGACAAATCAATAATACATGTAGAAGCGAACAATGGTCAGAACCA	682
Oy		501	ProArgCysLeuAspProCysValAlaSerGluLysMet-AsnLysasn--AsnI	519
Db		683	CCAAAAGCTTAGATCCATGTGTATATACCAAGAAATTCATGCAACAATTATTAACAT	742
Oy		519	egLnLeuLys--TriprArgAsnaSpGLyLysLeuTyala-LysThrGlyAspAlaValG	538
Db		743	AAATTTCAAGTGAGGACAAACAAACAAAMAACCTTATTTCACGAAACGGGTACATAGTTG	802
Oy		538	Iupha	539
Db		803	AATTC	807
RESULT 7				
BGS69423				
LOCUS		858 bp	mRNA	linear EST 10-APR-2001
DEFINITION		602588766F1 NIH_MGC_76	Homo sapiens CDNA clone IMAGE:4722761 5'	
ACCESSION		BGS69423		
VERSION		BGS69423.1	GI:13577076	
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE		1 (bases 1 to 858)		
JOURNAL		NIH-MGC htp://mgc.nci.nih.gov/.		
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. CDNA Library Preparation: CLONTECH Laboratories, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LNLN at: http://image.lnl.gov Plate: LICML1581 row: C column: 18 High quality sequence stop: 758. Location/Qualifiers 1..858 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4722761" /clone_lib="NIH MGC.76" /lab_host="DH10B (TI phage-resistant)" /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1; Sfil (ggcgccctggccc); Site:2: Sfil (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGGCAGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC library."		
BASE COUNT		276 a	181 c	157 g 244 t
ORIGIN				
Alignment Scores:		1.58e-75	Length:	858
Pred. No.:				

Score:	797.50	Matches:	183
Percent Similarity:	36.44%	Conservative:	24
Best Local Similarity:	32.22%	Mismatches:	56
Query Match:	25.05%	Indels:	306
DB:	12	Gaps:	2
US-09-805-337A-2 (1-569) x BG569423 (1-858)			
QY	1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20		
DB	69 ATGTGGCTGCTGGTCAGTGTATTCATTCACATCTCAGCATATTCCTCTTGGGGAGCAAGCA 128		
QY	21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTrpAspGluAspTyrAsn 40		
DB	129 ATGTTCTGTGATTTTCCAAAATAAACCATGTGATTCCTATATGATGATGAGAAAATATTAAG 188		
QY	41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluLysTrpAsnPheVal 60		
DB	189 CCATTTTCCCAAGTTCCTACAGGGGAGTTTCCTATTAATCTCCGTGAATATATTTTGTG 248		
QY	61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr 80		
DB	249 TCTCTCTCAATTCCTTTGGAGCTGCCATTAACCTGGCAGAAAGATGCTACCAACA 308		
QY	81 ProLysCysLeuArgNecCysSerPheProPheVal-LysAsnGlyHisSerGluSer 100		
DB	309 CCAAGTGTCTCAGACTGTGTCTTCTCTCTTTGTGTGAAACAGTCATTCGAATCTTC 368		
QY	100 rGlyLeuIleHisGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerIe 120		
DB	369 AGCAACAACACATCTGGAGGTGATACCTGACAAATTAATTTGCACACAGGATACAGACT 428		
QY	120 uGlnAsnAsnGlyLysAsnIleSerCysValGluArgIlyTrpSerThrProProIleCy 140		
DB	429 TCAAAACATGAGACACACATTTCACTGTAGACAGCGGCTGTCCACTCTCCCAATG 488		
QY	140 sSerPheThrLysGlyGlyCysHisValProIleGluGluAlaAsnValAspAlaGlnPr 160		
DB	489 C----- 489		
QY	160 oLysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuI 180		
DB	489 ----- 489		
QY	180 eArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnPheProTh 200		
DB	489 ----- 489		
QY	200 rCysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLy 220		
DB	489 ----- 489		
QY	220 sGluIleArgLysGluGluTyrGlyHisAsnGluValGluTrpAspCysAsnProAs 240		
DB	489 ----- 489		
QY	240 nPheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPr 260		
DB	489 ----- 489		
QY	260 oThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTrpGlyTyrVa 280		
DB	489 ----- 489		
QY	280 lGlnProSerValProProTyrGlnHisGlyLysSerValGluValAsnCysArgAsnGly 300		
DB	489 ----- 489		
QY	300 uTyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPr 320		
DB	489 ----- 489		
QY	320 oMetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysTh 340		

```

Db      489 ----- 489
QY      340 rleuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSe 360
Db      489 ----- 489
QY      360 rAspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAs 380
Db      489 ----- 489
QY      380 pCysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnIleGlnAs 400
Db      489 ----- 489
QY      400 mMetThrThrThrValAsnTyrGlnAspGlyLysValAlaValLeuCysLysGluAs 420
Db      489 ----- 489
QY      420 nTyrIleLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPr 440
Db      489 ----- 489
QY      440 oArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspTh 460
Db      490 -AGGTCACACTATTTCGACAGAAAATGTCGGCCCTCCACCTATTGACAAATGAGACAT 548
QY      460 rThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSe 480
Db      549 TACTTCATTCCTGCTGTCTGATATAGCTCTCAGCTTCATCATGTTGAGTACAGAGTCAGAGA 608
QY      480 rPheTyr-LysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluP 500
Db      609 CTGTGATCAAACTGATGAGGTGATCAATCAATCAATGAGAAAGGACAAATGCTGAGAAC 668
QY      500 rCPeArgCysLeuAspProCysValValSerGluGlnAsnMetAsnLysAsnAsnIleG 520
Db      669 CACCAAAATGCTTATGATTCATGTTATATCCCAAAATTTATGCAACAATATACCTTAA 728
QY      520 InLeuLysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheG 540
Db      729 AATTACAGTGGACAAAGCAACAAAGCTTATCC-AAGACAGTGCATATTATGACTTGG 787
QY      540 InCysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnG 560
Db      788 TTCGTAACCTCCGAAATCATATCCCAAAATCTCA-----TTCAATTGACGATGTTTGG 838
QY      560 lGlyLysPheGluTyrPro 566
Db      839 AATTGCAACTGTATTTCCCG 858

RESULT 8
AUI38301 853 bp mRNA linear EST 02-AUG-2002
LOCUS AUI38301 PLACel Homo sapiens cDNA clone PLACE1008277 5', mRNA
DEFINITION AUI38301.1 GI:10999822
ACCESSION AUI38301
VERSION AUI38301.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saiko,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

```

```

Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
FEATURES
source
1. 853
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1008277"
/tissue_type="placenta"
/issue="Vector"
/note="Vector"
BASE COUNT 297 a 152 c 185 g 214 t 5 others
ORIGIN
Alignment Scores:
Pred. No.: 5,45e-75 Length: 853
Score: 792.50 Matches: 143
Best Local Similarity: 72.84% Conservative: 34
Best Local Similarity: 58.85% Mismatches: 54
Query Match: 24.90% Indels: 12
DB: 9 Gaps: 3
US-09-805-337a-2 (1-569) x AUI38301 (1-853)
QY 226 GUTYRGYLYHISASNGLUVALYALGUTYRAspCysAsnProAsnPheIleAsnGly 245
Db 2 GAATATGACACAGCAAGCAAGTGTGAAATTTATTCGAAATTCGATTAATGAGGGA 61
QY 246 ProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuProThrCysValGluGln 265
Db 62 CCTAATTAATTCAGTGCTTATGATGAGAGTGCAGACATTACAGTGTGTGTGCGAG 121
QY 266 ValLysThrCysGlyTyrIleProGluLeuGluTyrGlyTyrValGlnProSerValPro 285
Db 122 GAGAGTACCTGCGAGATATACCTGAACTTGAACATGGCTGGCCAGCTTTCCTCCCT 181
QY 286 ProTyrGlnHisGluValSerValGluValAsnCysArgAsnGluTyrAlaMetIleGly 305
Db 182 CCTTATTAATGAGATTCAGTGCATTCATCAATTCGTCAGAAATCAATTAATGATGGA 241
QY 306 AsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuProMetCysValAlaThr 325
Db 242 CACAGATCAATTACGATTCATTCATGAGATGACCAAACTTCCCAAGTGTGCAATA 301
QY 326 HisGlnLeuLysArgCysLys-----IleAlaGlyValAsnIleLysThr 340
Db 302 GATTAACCTTAAGCAAGTGCMAATCATCAATTAATTAATTAATTAATTAATTAATAA 358
QY 344 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
Db 359 -----AACAGAGGAATTCGATCAATTAATTAATTAATTAATTAATTAATTAATTA 406
QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
Db 407 GGAAGAAAGAGATGATGATACACACAGTCGCAATTAATGAGAGTGGATGCCAAGTGAAC 466
QY 381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnIleGlnAsn 400
Db 467 TGCTCAATGCGCAAAATCAATTAATGAGTCCACCTCCACCTCAGATTCCTCAAT 526
QY 401 MetThrThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCys-LysGluAs 420
Db 527 ATGACACACACACTAATTAATGAGTGGATGAGAAAGATATCTTGTTCGCAAGAAANA 586
QY 420 nTyrIleLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPr 440
Db 587 TTATCTAATTATGCAAGAGCAAGAAATTAATGCAAGATGCAAGATGCAATGCAATGCC 646
QY 440 oArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspTh 460

```

```

Db      647 ACTGTGTGTGMAAAATTCCA---TGTTCACACCAACCTCAGATAGAACAGCGACCAT 703
QY      460 rThrsr 462
      704 TAAATCA 710

RESULT 9
BG564962
LOCUS   602583332p1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711903 5',
DEFINITION
mRNA sequence.
ACCESSION
BG564962
VERSION
BG564962.1 GI:13572615
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 912)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1552 row: o column: 08
High quality sequence stop: 713.

FEATURES
Source
Location/Qualifiers
1..912
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4711903"
/clone.lib="NIH_MGC_76"
/lab.host="DH10B (TI phage-resistant)"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccgcgcgc); Site_2: SfiI (ggcattacggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCATATGAGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGGCGCATG-3' (30)Bv-3' (where B = A,
C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

BASE COUNT
296 a 187 c 173 g 256 t

ORIGIN
Alignment Scores:
Pred. No.: 1,63e-74 Length: 912
Score: 788.50 Matches: 176
Percent Similarity: 35.86 Conservative: 18
Best Local Similarity: 32.53 Mismatches: 43
Query Match: 24.77 Indels: 304
DB: 12 Gaps: 2

US-09-805-337a-2 (1-569) x BG564962 (1-912)
QY      1 MetLeuLeuLeuPheSerValIleLeuIleSerTTPValSerThrValGlyGlyGly 20
Db      84 ATGTGCTCTCTGTCAGTGAATTAATCTAACCTCAGGATATCCCTCTGTGGGGGAGAGCA 143
QY      21 ThrLeuCyAspPheProLysIleHisIsglyPheLeuTyAspGluIuAspTyraAsn 40
Db      144 ATGTTCTGTGATTTTCCAAAATAAACCCTGGAATTCATATGATGATAAATAATTTAAG 203
QY      41 ProPheSerGlnValProThrGlyGluValPheTyTYrSerCysGluTyraAsnPheVal 60

```

```

Db      204 CCATTTCCTCAAGTTCCTACAGGAGGAGATTTCATTACTCTGTGAATATATTTTGG 263
QY      61 SerProSerLysSerPheThrPThrArgIleThrCysThrGluGlyTYrSerProThr 80
Db      264 TCTCCTTCAAAATCCCTTTGGACTCGCATACCTCCGAGAGAGATGTCACCAACA 323
QY      81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerCysLysSer 100
Db      324 CCAAGCTGTCTCAGACTGTGTCTTCTCTTTGTGGAAAATGTCATTTCTGAATCTTCA 383
QY      101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTYrSerLeu 120
Db      384 GCACAAACACATCTCGAAGGTGATCTGTACAAATATTATTCACACAGGATACAGACTT 443
QY      121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTYrPheThrProProIleCys 140
Db      444 CAAACCAATGAGAACACATTTCAATGATGACGAGCGCTGTCCACGCTCCCAAAATGC 503
QY      141 SerPheThrLysGlyGlyCysHisValProIleLeuGluValAsnValAspAlaGlnPro 160
Db      503 -----
QY      161 LysLysGluSerTyLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
Db      503 -----
QY      181 ArgValGlySerAspSerValGlnCysTyGlnPheGlyTYrPheProAsnProThr 200
Db      503 -----
QY      201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLys 220
Db      503 -----
QY      221 GluIleArgLysGluGluTyGlyHisAsnGluValGluTyAspCysAsnProAsn 240
Db      503 -----
QY      241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTYrThrLeuPro 260
Db      503 -----
QY      261 ThrCysValGluGlnValLysThrCysGlyTYrIleProGluLeuGluTYrGlyTYrVal 280
Db      503 -----
QY      281 GlnProSerValProProTYrGlnHisGlyValSerValGluValAspCysArgAsnGlu 300
Db      503 -----
QY      301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTYrThrGluLeuPro 320
Db      503 -----
QY      321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
Db      503 -----
QY      341 LeuLeuLysLeuSerClyLysGluPheAsnHisAsnSerArgIleArgTYrArgCysSer 360
Db      503 -----
QY      361 AspIlePheArgTYrArgHisSerValCysIleAsnGlyLysTYrPheAsnProGluValAsp 380
Db      503 -----
QY      381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
Db      503 -----
QY      401 MetThrThrThrValAsnTyGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420

```

Db	503	-----	503
Qy	421	TyrIeuLeuPProGluAlaValGluIleValCysLysAspGlyArgTyrPglInSerIeuPro	440
Db	503	-----	503
Qy	441	ArgCysValGluSerThrIaLysCysGlyProProSerIleAsnAsnGlyAspThr	460
Db	504	AGGTCACCTATTTCGCGAAAAATGTCGGCCCTCCACCTATTGACAAATGAGACATT	563
Qy	461	ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer	480
Db	564	ACTCATTCCTCGTGTGTAGTATATGCTCCAGGTTATCAGTTGAGTACCACTGCAGAC	623
Qy	481	PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrPserGluP	500
Db	624	TTGTTCACACTTGAGCGTGAACAAATCAATGATGTAAGAACCGACAACTGTCAGAAC	683
Qy	500	oProArgCysIeuAspProCysValValSerGluLysn---MetAnLysAsnAsnI	519
Db	684	ACCAAAATGCTTAGTCATGTTATTCACACAGAAATATTGACACAAATTTACCAT	743
Qy	519	eGluIeu---LysTyrPArgAsnAspLysLysLeuTyrAlaLysThrGlyAspAlaVal	537
Db	744	TAAATTTAAAGTGAACAAACACAAAGCTTTATTTCAGCACCGGTGACATAGTT	800
RESULT 10			
BG533258			
LOCUS	BG533258	786 bp	mRNA linear EST 03-APR-2001
DEFINITION	6018605232 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4072108 5',		
ACCESSION	BG533258		
VERSION	BG533258.1	GI:13524798	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhina; Homalidae; Homo.		
TITLE	1 (bases 1 to 786)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@rs-remail.nih.gov Tissue Procurement: CLONETECH Laboratories, Inc. cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.llnl.gov Plate: L1CM918 row: m column: 05 High quality sequence stop: 762. Location/Qualifiers 1. 786 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image:4072108" /lab_host="NIH_MGC_76" /lab_host="DH10B (TI phage-resistant)" /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccgcgcgc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were in cloning as follows: 5' adaptor sequence: 5'-CAGCGCATTAATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGAGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."		
BASE COUNT	264 a	157 c	146 g 219 t

Alignment Scores:			
Pred. No.:	2,76e-74	Length:	786
Score:	785.50	Matches:	172
Percent Similarity:	36.65%	Conservative:	23
Best Local Similarity:	32.33%	Mismatches:	36
Query Match:	24.68%	Indels:	302
DB:	12	Gaps:	2
US-09-805-337A-2 (1-569) x BG533258 (1-786)			
QY	1 MetLeuLeuLeuPheSerValIleLeuIleSerThrPValSerThrValGIGlYglUcYl		20
DB	84 ATGGGCTCCTGGTCAGTGAATTAATCTACGGATATCTCTGTTGGGGAGAAACA		143
QY	21 ThrLeuCysAspPheProLysIleHisGlyPheLeuYrAspGluAspTyrAsn		40
DB	144 ATGTTCTGATTTTCCAAATAATAAACCATGGAATTATATGATGAAGAAATAATATAG		203
QY	41 ProPheSerGlnValProThrGlyGluValIlePheTYrTYrSerCysGluTYrAsn		60
DB	204 CCATTTTCCCAAGTCTCTACAGGGGAAGTTTCTATTACTCTGTGAATATAATTTGTG		263
QY	61 SerProSerIysSerPheThrPThrArgIleThrCysThnGluGluTYrPThr		80
DB	264 TCTCTTCAAAATCCTTTTGAGCTGCATTAACGTGCGAAGAAAGATGGTCAACACA		323
QY	81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer		100
DB	324 CCAAGAGTCTCAGACGTGTTCTTCTTTTGAGAAATGTCATTTGATTCATCTCA		383
QY	101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTYrSerLeu		120
DB	384 GGACAAACACATCTGGAGAGTGATTACTGACAAATTAATTGGCAACACAGATACAGACTT		443
QY	121 GlnAsnAsnGlyLysAsnIleSerCysValGluATrGlyTYrPThrProProIleCys		140
DB	444 CAAACATATGAGAACACACATTTTCATGTGAGAACGGGTGTCACATCTCCCAAAATGC		503
QY	141 SerPheThrLysGlyLysCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro		160
DB	503 -----		503
QY	161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle		180
DB	503 -----		503
QY	181 ArgValGlySerAspSerValGlnCysTYrGlnPheGlyTYrPThrProAsnPheProThr		200
DB	503 -----		503
QY	201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLys		220
DB	503 -----		503
QY	221 GluIleArgLysGluGluTyrGlyHisAsnGluValValGluTYrAspCysAsnProAsn		240
DB	503 -----		503
QY	241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro		260
DB	503 -----		503
QY	261 ThrCysValGluGlnValLysThrCysGlyTYrIleProGluLeuGluTYrGlyTYrVal		280
DB	503 -----		503
QY	281 GlnProSerValProProTYrGlnHisGlyLysValGluValAsnCysArgAsnGlu		300
DB	503 -----		503
QY	301 TyrAlaMetIleGlyLysAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro		320
DB	503 -----		503

FEATURES	LOCATION/Qualifiers
OY	321 MetCysValAlaThrTrhHisGlnLeuLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
Db	503 ----- 503
OY	341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
Db	503 ----- 503
OY	361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
Db	503 ----- 503
OY	381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
Db	503 ----- 503
OY	401 MetThrThrThrValAsnTyrGlnAspCysLysGluLysValAlaValLeuCysLysGluAsn 420
Db	503 ----- 503
OY	421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrPcinSerLeuPro 440
Db	503 ----- 503
OY	441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460
Db	504 AGGCCCACTATTTCGCAGAAAAATGTGGCCCCCTCCACCTATTGGACAAATGGAGACATT 563
OY	461 ThiserPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
Db	564 ACTTCATTCCTGGTGTGAGTATATGTCACAGGTTTCATCAGTTGAGTACCACTGCCAGAAC 623
OY	481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500
Db	624 TTGTATCAACTGGAGGTATACATCAA-ATAACATGTATCAAAACGCAACATGTCAGAACCA 682
OY	501 ProArgCysLeuAspProCysValValSerGluGlnAsnMetAsnLysAsnAsnIleGln 520
Db	683 CCAAAATCCTTAGATCCATGTGTATATACACAGAAA---TTATGGAAAAAATATTAACATTA 739
OY	521 LeuLysTrpArgAsnAspCysLysLeuTyrAlaLys 532
Db	740 ATTAAGTGGCAACAACAAACAAACCTTATTCAGAA 775
RESULT 11	
LOCUS	BG533248 839 bp mRNA linear EST 03-APR-2001
DEFINITION	601860612P2 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4071962 5',
ACCESSION	BG533248
VERSION	BG533248.1 GI:13524788
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strusberg, Ph.D. Email: rstrus@rmail.nih.gov Tissue Procurement: CLONETECH Laboratories, Inc. cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: L10C918 row: g column: 03 High quality sequence stop: 838.

```

Source
1. 839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1 IMAGE:4071962"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgagctcgagc); Site_2: SfiI (ggccatctagc); 5' am
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCATATGGCC-3' and 3' adaptor sequence
5'-ATTCTAGAGCGCCGAGCGCCGCGCATATG-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.8
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT
273 a 176 c 155 g 235 t

```

Alignment Scores:			
Pred. No.:	1,2e-73	Length:	839
Score:	780.00	Matches:	175
Percent Similarity:	36.41%	Conservative:	22
Best Local Similarity:	32.35%	Mismatches:	41
Query Match:	24.51%	Indels:	304
DB:	12	Gaps:	2
US-09-805-337A-2 (1-569) x BG533248 (1-839)			
QY	1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly	20	
DB	66 ATGTGCTCTCGTCAGTGTAACTTCTAATCTCAGGATATCTCTGTGGGAGAGCA	125	
QY	21 ThrLeuCyAspPheProLysIleHSHSGlyPheLeuTyAspGluLyspTyAsn	40	
DB	126 ATGTTCTGTGATTTCCAAATAAATTAACCATGGAATTCCTATGATGAGAAATAATTAAG	185	
QY	41 ProPheSerGlnValProThrGlyGluValPheTyTrpSerCysGlyTrpAsnPheVal	60	
DB	186 CCATTTTCCAGAGTCTCTACAGGCAAGTTTCTATTACTCTGTGAATTAATTTTGTG	245	
QY	61 SerProSerLysSerPheTrpThrArgIleThrCysFhrGluGluGlyTrpSerProthr	80	
DB	246 TCTCCTTCAAAATCCTTTTGAGCTGCCATACGTCGAGAGAGAGAGATGTGCACCAACA	305	
QY	81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer	100	
DB	306 CCAAAAGTGTCCAGCATGTGTTCTTCTTTTGTGGAATAATGCTCATTCGAACTTCA	365	
QY	101 GlyLeuIleHisLeuGluGluLysAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu	120	
DB	366 GGACAAACACATCTGGAGAGGTATACTGTACAAATTAATTTGCACACAGGATACACACTT	425	
QY	121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys	140	
DB	426 CAAACAACTGAGACAACACATTTTCATGTGTAGAAGAGGGGGTGGTCCATCCCAATGC	485	
QY	141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro	160	
DB	485 -----	485	
QY	161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle	180	
DB	485 -----	485	
QY	181 ArgValGlySerAspSerValGlnCysTyArgInPheGlyTrpSerProAsnPheProThr	200	
DB	485 -----	485	
QY	201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyValLys	220	
DB	485 -----	485	

QY	221	GIUILEARGLGLUGLUTGRLYHSAASNGUVALVALGLUTYRASPQSYASNP	240
Db	485	-----	485
QY	241	PhelIleAsnGLYProLysLysIleGlnCysValAspGLYUtrPHrThrLeuPro	260
Db	485	-----	485
QY	261	ThrcysValGLUGlnValLysThrcysGLYTrIleProGlnLeuGLUTyrGLYVal	280
Db	485	-----	485
QY	281	GlnProSerValProProTyrGlnHscGLYAlSerValGluValAsnCysArgAsnGLU	300
Db	485	-----	485
QY	301	TyrAlaMetIleGLYAsnAsnMetIleThrcysIleAsnGLYIleTrpThrGluLeuPro	320
Db	485	-----	485
QY	321	MetCysValAlaThrHscGlnLeuLysArgCysLysIleAlaGLYAlaAsnIleLysThr	340
Db	485	-----	485
QY	341	LeuLeuLysLeuSerGLYLysGluPheAsnHISAsnSerArgIleArgTyrArgCysSer	360
Db	485	-----	485
QY	361	AspIlePheArgTyrArgHisSerValCysIleAsnGLYIleSerTrpAsnProGluValAsp	380
Db	485	-----	485
QY	381	CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn	400
Db	485	-----	485
QY	401	MetThrThrValAsnTyrGlnAspGLYGLUValAlaValLeuCysLysGluAsn	420
Db	485	-----	485
QY	421	TyrLeuLeuProGluAlaLysGluIleValCysLysAspGLYArgTrpGlnSerLeuPro	440
Db	485	-----	485
QY	441	ArgCysValGluSerThrAlaTyrCysGLYProProProSerIleAsnAsnGLYAspThr	460
Db	486	AGGTCACCTATTTCGCAAAAAAAGTGGGCCCCCTCCACCTATGACATGAGAGACATT	545
QY	461	ThSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer	480
Db	546	ACTTATTCCTGCTGTGTCACTATATGCTCCAGGTTCACTGAGTACCAAGTCCGGAAC	605
QY	481	PheTyrLysLeuGlnGLYSerValThrVal-ThrcysArgAsnLysGlnTrpSerGluBr	500
Db	606	TTGCTATCAACTTGACAGGTACACATCAATCAATAAATCAATGTAACAAACGACATGGTCAGAAC	665
QY	500	Pro-ArgCysLeuAspProCysValValSerGlnGluAsnMet--AsnLysAsnAsnI	519
Db	666	ACCAAAAAAGCTTAATCCATGTGTAAATTCACAAAGAAATTTCTGGGCAACAT-TATACA	724
QY	519	IleGlnLeuLysTrpPArgAsnAspGLYLysLeuTyrAlaLysThrGlyAspAlaValGlu	538
Db	725	TAAATTAAAGTGGACAAACAAACAAAGTTTCATTCCAGAAACAGGTGACATAGTTGAC	783

RESULT 12

LOCUS BG286384

DEFINITION 603383156P1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4500732 5',

ACCESSION BG286384

VERSION BG286384.1 GI:13039218

KEYWORDS EST.

SOURCE human.

ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	1 (bases 1 to 790)			
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabps-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov plate: LHAM10366 row: p column: 13 High quality sequence stop: 661. Location/Qualifiers 1..790 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_1b="NIH_MGC_93" /clone_1b="NIH_MGC_93" /tissue_type="transitional cell papilloma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: bladder; Vector: pCMV-Sport6; Site:1; NotI; Site:2; SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."			
BASE COUNT	274 a 146 c 166 g 204 t			
ORIGIN				
Alignment Scores:				
Score, No.:	2,31e-73	Length:	790	
Percent Similarity:	777.00	Matches:	138	
Best Local Similarity:	72.57%	Conservative:	34	
Query Match:	58.23%	Mismatches:	65	
DB:	12	Indels:	0	
		Gaps:	0	
US-09-805-337A-2 (1-569) x BG286384 (1-790)				
OY	192	PheGlyTyrSerProAsnPhcProThrCysLysGlyGlnValArgSerCysGlyProPro	211	
Db	5	TTTTGGATTGCTCTCTGACCTCCCAATATGTAAAGAGCAAGTCAATCATGTGGTCCACCT	64	
OY	212	ProGlnLeuSerAsnGlyGlnValLysGluIleArgLysGlnGluTyrGlyHisAsnGlu	231	
Db	65	CTGTAACTCTCATATGGGAATGTTAAAGGAAAAACGAAAGAAATATATGGACACAGTAA	124	
OY	232	ValValGluTyrAspCysAsnProAsnPhelIleAsnGlyProLysLysIleGlnCys	251	
Db	125	GTGGTGAAGATATTATGGCAATCTCTAATGTAAGAACGACCAATAAATAATCAATCT	184	
OY	252	ValAspGlyLysIlePheThrIleuPProThrCysValGlnGlnValLysThrCysGlyTyr	277	
Db	185	GTGTATGAGAGAGGCAACTTTTACACAGTGATTTGGAGAGAGTAACTGTGGCAAT	244	
OY	272	IleProGluLeuGluTyrGlyTyrValGlnProSerValProProTyrGlnHisGlyVal	291	
Db	245	ATACCTGAACCTTGAACATGGCTGGCGCCAGCTTCTGCCCTTATTTACTATGAGAT	304	
OY	292	SerValGluValAsnCysArgAsnGlyTyrAlaMetIleGlyAsnAsnMetIleThrCys	311	
Db	305	TCAGTGAATTCATTCAATGCTCAGAAATCATTTTAAAGATGATTGGACACGATCAATACCTGT	364	
OY	312	IleAsnGlyIleIlePheThrGluLeuPProMetCysValAlaIleHisGlnLeuLysArgCys	333	
Db	365	ATTTCATGGAAGTATGAGACCCCAACTTCCCAAGTGTGGCAATGATTAACATTAAAGAGTGC	422	
OY	332	LysIleAlaGluValAsnIleLysThrIleuLeuLysLeuSerGlyLysGluPheAsnHis	351	

Db	425	AAATCATCAAAATTTAAATTATCTTGAGGAGACATTTAAAAACAGAAAGCATTCGATCAT	484
QY	352	AsnSerAlaGleAaGtYrAaGcYSerAspIlePheArGtYrAaGHisSerValCysIle	371
Db	485	AATCTCAACAAATAGGTACAGATGTAGAGAAAGAAAGAGATGAAACACAGCTGTCATA	544
QY	372	AsnGlySTrPAsnProGluValAspCysThrGluLysAspGluInPheCysProPro	391
Db	545	AATGGAAGATGGATCCGAAGTAACTGGTCATCGACCAAAATTCATTTATGCCACCT	604
QY	392	ProProGluInIleProAsnAlaGlnAsnMetThrThrThrValAsnTyrglnAspGly	411
Db	605	CCACCTCGATGATCCCAATTCACAAATTTGTACACACACATGAAATTCATCGGATGAGAA	664
QY	412	LysValAlaValLeuGlyGlnAsnTyrglnLeuProGluAlaLysGlu	428
Db	665	CACAGTATCTGCTCTTGCCAGGAAATATATCTAATTACGAGAGAGAA	715
RESULT 13			
LOCUS	AV682232	815 bp	mRNA linear EST 16-JAN-2002
DEFINITION	AV682232	GKB Homo sapiens cDNA clone GKBAG04 5', mRNA sequence.	
ACCESSION	AV682232		
VERSION	AV682232.1	GI:10284095	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 815)		
AUTHORS	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.		
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver		
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)		
COMMENT	21625106 Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.		
FEATURES	Location/Qualifiers		
source	1..815		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="GKBAG04"		
	/clone_lib="GKB"		
	/tissue.type="hepatocellular carcinoma"		
	/dev.stage="Adult"		
	/lab_host="SOLR"		
	/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	252 a 166 c 173 g 223 t	1	others
ORIGIN			
Alignment Scores:			
Pred. No.:	2 42e-73	Length:	815
Score:	777.00	Matches:	152
Percent Similarity:	66.27%	Conservative:	17
Best Local Similarity:	59.61%	Mismatches:	79
Query Match:	24.41%	Indels:	7
DB:	10	Gaps:	3
QY	US-09-805-337A-2 (1-569) x AV682232 (1-815)		
QY	1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGluGly	20	

Db	62	ATGAGGCTCCGAGTCAGTGTATTTCTATCTCAGGATATTCCTCTGTGGGAGAAAGCA	121
QY	21	ThrlcucysasphrpoluylsHisHisGlyPheLeuTyraSpGluAspTyraSn	40
Db	122	ACATTTGTGATTTTCCAAAATATACCATGGAATCTATATCATGATGAAGAAATATTAAG	181
QY	41	ProPheSerGluValProThrGluGluValPheValTyTrSerCysGluTyraAsPheVal	60
Db	182	CCATTTTCCAGGTTCTACAGGGGAAGTTTGTATTACTCTCTGTGATTAATATTTTGTG	241
QY	61	SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr	80
Db	242	TCCTCTCAAAATCATTTTGTGACCTCGCATACATGACAGAGAAGAAGATGGTCCACAA	301
QY	81	ProLysCysLeuAlaGmetCysSerSerPheProPheValLysAsnGlyHisSerGluSer	100
Db	302	CCAAAGGTCTCAACACGTGCTTTCTTTCTTTGTGAAAATGTCATTCGAATCTTCA	361
QY	101	GlyLeuIleHisLeuGluGluLysAspThrValIsnIleLeuLysAsnThrGlyTySerLeu	120
Db	362	GGACAACAACATCTCGAAGGTGATACCTGTCAATTTATTTCCACACAGATACAGACTT	421
QY	121	GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProLecys	140
Db	422	CAAAACATGAGAACACATTTTCATGTGTAGAACGGGGCTGTGCACCCCTCCAAATGC	481
QY	141	SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsn-----ValAsp	157
Db	482	AGGTCCATGACACTCTCCTGTGTGAATCCGGCCACAGTACAAAATGCTTATATAGTCTCG	541
QY	158	AlaGlnProLysLysGluSerTyTrLysValGlyAspValLeuLysPheSerCysArgLys	177
Db	542	AGACAGATGAGTAA-----TATCCATCTGTGAGAGATGACTTATCATATGTAGAGAC	595
QY	178	AsnLeuIleArgValGlySerAspSerValGlnCysTyTrGlnPheGlyTrpSerProAsn	197
Db	596	CCTTATGAAATGTTTGGGGATGAAGAAATGATGCTTAAATGGAACATGAGCGAACA	655
QY	198	PheProThrCysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGly	217
Db	656	---CCTCATGCAAAAGATTCTACGGGAAATCTGGGCCCTTNCACCATTTGACAATGG	712
QY	218	GluValLysGluIleArgLysGluGluTyTrGlyHisAsnGluValValGluTyraSpCy	237
Db	713	GACATTACTTATCTCCCGCTCAGAAATCTCCAGCTTCAAGATGAGTTCAGTCCCATGG	772
QY	237	sAsnProAsnPhelIleAsnGlyProLysLysIleGlnCys	251
Db	773	CAAGACTGTATCACTTGGAAGGACACAGATACCATGT	815
RESULT 14			
LOCUS	AV688058	761 bp	mRNA linear EST 16-JAN-2002
DEFINITION	AV688058	GKC Homo sapiens cDNA clone GCKDM03 5', mRNA sequence.	
ACCESSION	AV688058		
VERSION	AV688058.1	GI:10289921	
KEYWORDS	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,G., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001) 21625106		
JOURNAL	CONTACT: Zeguang Han		
MEDLINE			
COMMENT			

QY 500 OProArGysLeuAspProCysValValSerGIuGIuAsnMetAsnLysAsnAsnIleGI 520
 Db 669 AGCAAAATGCTTAGATCCATGCTGAATATACACAAG--AATTATGGAATAATATACATPAA 726
 QY 520 nLeuLys 522
 Db 727 ATTAAAG 733

Search completed: July 23, 2003, 09:39:43
 Job time : 2209 secs

THIS PAGE BLANK (USPTO)